DOCUMENT RESUME

ED 310 368 CS 009 773

AUTHOR	Anderson, Thomas H.; Huang, Shang-Cheng Chiu
TITLE	On Using Concept Maps To Assess the Comprehension Effects of Reading Expository Text. Technical Report No. 483.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Aug 89
CONTRACT	OEG-0087-C1001
NOTE	35p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Concept Formation; Concept Teaching; Grade 8; Junior High Schools; *Reading Comprehension; Reading Instruction; Reading Research; *Reading Strategies; *Reading Tests; Student Attitudes
IDENTIFIERS	*Concept Mapping; Concept Maps; *Expository Text

ABSTRACT

A study examined: (1) the feasibility of using concept maps as a measure of content achievement; and (2) whether these measures were sensitive to knowledge learned after reading expository text. Subjects, 131 eighth-grade students, were taught a mapping technique which required them to analyze ideas into propositions and then arrange them into concept maps. Some students were assigned to one of two groups: one group read a content passage; the other read the passage and saw slides. The remaining students were in a control group which received no instruction. All students took a mapping test, a short answer "classroom-type" test, and an attitude questionnaire. Results showed that most students from all ability groups learned to use the mapping technique. Scores from the mapping test were more sensitive to the increases in knowledge gained by reading and seeing the slide presentation than were those from the short-answer test. In addition, the mapping test scores correlated with other indices of schooling progress such as classroom grades and standardized measures of achievement. Although the diagnostic value of the tests seemed to be substantive and could be helpful to teachers and students as they attempt to determine what they do and do not understand, students did not express a positive attitude toward the mapping tests, as indexed by the attitude questionnaire. (Five figures and seven tables of data are included, and 25 references are attached.) (Author/SR)

CENTER FOR THE STUDY OF READING

Technical Report No. 483

ON USING CONCEPT MAPS TO ASSESS THE COMPREHENSION EFFECTS OF READING EXPOSITORY TEXT

Thomas H. Anderson
University of Illinois at Urbana-Champaign

Shang-Cheng Chiu Huang
National Taiwan University

August 1989

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. OEG 0087-C1001, with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

Abstract

Data were collected from eighth graders to determine the feasibility of using concept maps as a measure of content achievement and whether these measures were sensitive to knowledge learned after reading expository text. All students were taught a mapping technique which required them to analyze ideas into propositions and then arrange them into concept maps. Then, some students were assigned into two groups which either read a passage about the processes of green leaves, or they read this passage and saw related slides. The remaining students were in the control group and received no instruction about the processes of green leaves. All students took a mapping test, a short answer "classroom-type" test and an attitude questionnaire. Results showed that most students from all ability groups learned to use the mapping technique. In addition, those students who received an extra session of practice learned to map better than those who did not. Scores from the mapping test were more sensitive to the increases in knowledge gained by reading and seeing the slide presentation than were those from the short-answer test. In addition, the mapping test scores correlated with other indices of schooling progress such as classroom grades and standardized measures of achievement. Although the diagnostic value of the tests seemed to be substantive and could be helpful to teachers and students as they attempt to determine what they do and do not understand, students did not express a positive attitude toward using mapping tests in schools, as indexed by the attitude questionnaire.

ON USING CONCEPT MAPS TO ASSESS THE COMPREHENSION EFFECTS OF READING EXPOSITORY T? XT

Introduction

The two main goals of this project were to determine (a) the feasibility of using concept maps as a measure of content achievement, and (b) how sensitive an index of learning they are after students read expository text. While other researchers have striven toward these goals, we feel that the surface of possibilities has just been scratched.

The use of a mapping or diagramming technique to represent meaning implies that a decision has been made about how to model memory or cognitive structure. While several models may represent the same aspect of content knowledge, a good model must mirror faithfully the most important aspects of a knowledge structure. Thus, a good model of knowledge allows people to examine and reach similar conclusions about the structure of knowledge by looking only at the representations (Rumelhart & Norman, 1985). Other characteristics of a good model, or theory, include completeness, specificity, generality, parsimony, and plausibility (Sternberg, 1985). Few models meet all these requirements.

With this in mind we opted to experiment with the "semantic network model" to build concept maps in somewhat the same way that Dansereau and his colleagues (1979) and Champagne and her colleagues (1978, 1981) have for several years. Our rational for this decision follows.

Theoretically, semantic networks are used here as a generic term for a class of models involving nodes (concepts) interrelated by labeled links (relations). More specifically, "The network can be considered to contain many mini-networks of propositions. And the propositions, not the single nodes or associations, are the units of factual knowledge" (Klatzky, 19?0, p. 197).

The notion of a semantic network was first introduced by Quillian (1966) and later by Collins & Quillian (1969). They proposed a model (TLC, for Teachable Language Comprehender) to account for the comprehension and use of language in a natural way. In this model, knowledge structure in human long-term memory was represented by a directed, labeled, and hierarchical graph structure composed of a set of nodes interrelated by various kinds of associative links. Their model also had the property of <u>inheritance</u>--the assumption that whatever characteristics were true about a concept at one level were true at all lower levels in the structure.

Although the TLC model was generally accepted many inadequacies restricted its use. For instance, the set of link types was insufficient and limited (e.g., Brachman, 1977; Cohen, 1983). The set consisted of "isa," "hasa," and "can." Other links involving conjunctive disjunctive, spatial, temporal, causal, possessive and comparative relations were not presented (Cohen, 1983).

The hierarchical characteristic of the Collins and Quillian (1969) structures has also been questioned because much factual knowledge does not appear to be so organized, that is, lateral relationships greatly outnumber vertical relationships (Cohen, 1983). Thus, most spatial, temporal, causal, and comparative relations do not seem to need a top-to-bottom, general-to-specific structure in the form of a hierarchy. Rather, they appear to be lateral relationships. For example, there is no apparent vertical relationship between "the pupa stage is before the adult stage" and "cotton spinning results in thread." However, hierarchical networks can be viewed as a special case or part of a holistic semantic network.

Based on Quillian's (1966, 1968) and Rumelhart, Lindsay and Norman's (19720 models, Holley, Dansereau and their colleagues (e.g., Dansereau, Collins, McDonald, Holley, Garland, Diekhoff, & Evans, 1979; Holley, 1979; Holley & Dansereau, 1984) developed "Networking," an abstract, content-independent representational system. In Networking, students are trained to transform prose into hierarchically organized node-link diagrams using a set of six named links. The nodes involve concepts, and the links specify the relationships between these concepts. In addition, these six named links are classified into three types of representational structures: (a) hierarchies (type-part), (b) chains (lines of reasoning, temporal orderings, causal sequences), and (c) clusters (characteristics-definitionsanalogies). Holley and Dansereau (1984) report the effectiveness of teaching undergraduate students to Network college level text as a studying aid. Their conclusions support the idea that the process of Networking facilitates the retention of textbook material.

A related propositional structured network (the Concepts Structure Analysis Technique, or ConSAT) was developed by Champagne and her associates (1978, 1981). ConSAT was developed to function as an assessment technique, rather than as a studying technique. It was an individually administered task where fifth-grade subjects were instructed to sort a stack of 13 cards (each card contained an earth science word) into two substacks. One stack of words was those that the subject recognized, and the other was those that the subject did not recognize. Then, the subjects were asked to arrange the recognized words into related sets and point out the relationships. Finally, the subjects were allowed to go through the unrecognized words and attempt to fit them into the structure already produced.

The researchers developed a scoring scheme based on the attributes of a prototypic or "master" structure produced by subject-matter experts in earth science. The scheme worked like, first, class inclusions (Igneous is a class of rock; Granite is in the class of igneous) and transformation relations (Igneous rock forms when lava and magma cool) were identified by experts as the basic attributes of the structure of geology concepts used with fifth graders. Further, a graph was built including both hierarchical class-inclusion and cyclical transformation relations. This graph was reported to be as "economical" as possible, in terms of representing this set of geological concepts and their relationships.

Two working hypotheses were developed to validate the ConSAT task: (a) changes in ConSAT Scores could detect differences between students' pre-instruction and post-instruction representations of geology concepts, and (b) the subjects' post-instruction representations would be more congruent with the standard structure than those of pre-instruction representations. The results showed that both hypotheses were confirmed (Champagne, et al., 1981).

Concerning the usability of this technique, it can be easily administered but is terribly time consuming since it requires card sorting. Since it is administered individually, however, it provides abundant information about the subjects' understanding of the concepts and their relationships, as well as the subjects' thinking processes. Also, scoring students' graphs seemed a bit cumbersome since each graph was judged relative to six dimensions (with several levels in each dimension). The differences among the dimensions were difficult to distinguish. Champagne, et al., (1981) reported an interrater reliability of only 0.70 until the scoring scheme was revised. Then, they reported a more respectable index, 0.93.

The results of the ConSAT investigations were encouraging in that the ConSAT scores were sensitive to classroom instruction; however, we wanted to try to overcome some of the scoring and useability problems.

We also considered other techniques which have been investigated for their psychometric properties, such as Idea Mapping (Armbruster & Anderson, 1984; Surber, 1984), and Concept Mapping (Novak, Gowin & Johnson, 1985). Each of these appeared to be more constraining and specialized than we thought was necessary, and/or the scoring procedures were too cumbersome. Therefore, we decided to build on the work by Champagne, et.al. (1981) and Holley and Dansereau (1984), but to modify it so that it could be administered to a group of middle school students.

To meet the goals mentioned earlier, we set these criteria prior to the development of and experimentation with this assessment technique.

Criterion 1. The "technology" of taking the test should be simple enough so that all students can learn to do it, even the slower and less-motivated ones.

Criterion 2. The technique must be sensitive to reading/instruction. That is, an increase in students' test scores should be concomitant with an increase in students' knowledge.

Criterion 3. Students should be able to score higher on the test when a text containing information being tested is available at testing time.

Criterion 4. Performance on the test should be comparable to (correlate with) performance on other acceptable testing and grading techniques.

Criterion 5. The procedures and products of the mapping technique should be consistent with the instructional goals and practices of the school, and the integrity of the content area.

Method

The sequence of development activities followed this general plan. First, a content topic (the processes and functions of a leaf in green plants) was chosen and topic experts were located within the university community. In concert with those experts, a master map or knowledge structure was constructed that represented the ideas about green leaves that middle school students were likely to study. Then, a series of instructional materials was designed that would be used to teach students the meaning of the symbols used in the master map, and the procedures for constructing a map. Next, a master text was written that presented a prose version of the ideas represented in the master map. Finally, a set of slides was collected and a script prepared that explained the relationships among the ideas about green plants.

While this study was one integrated project, it will be described in two phases: (a) one in which students were trained to use the technique, a response to Criterion 1, and (b) one in which the characteristics and effects of the technique were investigated, a response to Criteria 2 through 5.

Phase I: Determining the Effects of Training on Learning to Map

Students

The students were eighth graders from two junior high schools. Seven intact classes, with a total of 170 students, were selected from the two schools in a small midwestern city. About one-half of the students were female and about 20% were minority students. Naturai attrition of students over the several days of instruction/testing and the unavailability of some standardized test information resulted in the total participation of 131 students.

Materials

Training materials. Two forms (Brief and Detailed) of the training materials were developed to help teach students how to take a mapping test. Form B consisted of Part I: definitions of the six groups of linkages accompanied by several examples, and Part II the step by step principles, procedures and examples of drawing a map. Form B required two instructional periods of about 50 minutes each. An example of the explanation and examples of a linkage is shown in Figure 1.

[Insert Figure 1 about here.]

Also, students were taught the following Principles and Procedures to be used in drawing a map:

Principles:

1. Use only the concepts and linkages given.
2. Each concept can be used only once.
3. Any linkage can be used more than once.
4. You do not need to use all linkages.
5. Arrows can be in any direction. They can connect any pair of concepts.

Procedures:

1. Read the test topic carefully.
2. Review all the concepts and think about the meaning of each one.
3. Think about the possible relationships among the concepts.
4. Choose the concept that is the central or most important one.
5. Write the central concept in the middle of the map.

6. Look at the list of possible relationships from Step 3. Find the concepts most closely related to the central concept.
7. Write the remaining concepts on the map. Place each concept close to related concepts which already appear on the map.
8. Draw arrows (links) between related concepts. Think about the direction of each arrow.
9. Label the arrows (links) using the relationships from Step 3.

Each of these Principles and Procedures was explained carefully and examples were provided as the experimenters/instructors illustrated each step as they made a map.

Form D (detailed training) contained the same materials as in Form B, plus abundant practice materials at the end of Parts I and II. The practice examples were based on content found in fifth to ninth grade life science textbooks. They included both concrete and abstract concepts which did not overlap with the content of the instructional materials that were used later in the instructional intervention and post-instructional tests. Form D required three instructional periods of about 50 minutes each.

Mapping Tests. All students took two mapping tests; one after Phase I to determine how well they had learned to map, and another after Phase II to determine how sensitive the test was to various instructional variables. We used our experience gained in an earlier study (Armbruster & Anderson 1984), from a pilot study using these procedures with 150 students in Taiwan, and from other research (Champagne et al., 1981: Cooke, Durso, & Schvaneveldt, 1986; Geeslin & Shavelson, 1975; Naveh-Benjamin, Lin, & Tucker, 1986; Novak et al., 1983) to help determine two important dimensions of a mapping test. These dimensions are the number and type of *concepts* and *linkages* to use. Other researchers used 10 to 25 concepts and about six linkages. We determined that approximately 15 concepts were appropriate. Many more than 15 seemed to create a bit of chaos on a single page, and fewer than 15 created a task that seemed to be too easy. In a similar vein, we supplied only those linkages that were necessary and sufficient to connect the 15 or so concepts. We did not include "filler" links to serve as distractors.

Content area experts played an active role in the development of these tests, especially in the process of deciding which concepts and linkages should be included. One of these experts, a plant pathologist, was particularly helpful because he used a similar mapping scheme in his lecturing/instructional program with undergraduate students.

The *concepts* included in the training test (Phase I), which was entitled, "Types of muscles and their functions," were: contraction, energy, heart muscle, involuntary muscles, movement, muscles, nerve impulse, outside stimulation, relaxation, skeletal muscles, smooth muscles, stomach, striations and voluntary muscles. The *linkages* used were: has a (is a part of), is characteristic of, is an example, results in, enables, is next to (takes place in).

The mapping test was administered in stages to determine how well students, (a) understood the *linkages*, and (b) could *draw* the map. On the linkage test, there were 14 items of the following type:

[enable]
1. Nerve impulses ---------> contraction.

is/are characteristic of
2. Smooth muscles ----------------------------> [the stomach].

Students wrote answers into the brackets [] using words from lists of concepts and linkages that were easily accessible.

The drawing stage of the Phase I mapping test was administered by giving students a so-called skeleton map and requesting that they write the 15 concept names, draw the arrows (including the arrow points) and label the arrows using linkage names. A skeleton map was an 8.5" by 14" piece of white paper with 44 evenly spaced, 3 cm long lines drawn on it.

Procedures

Students were randomly assigned to the brief or the detailed training groups. In the brief training group, students read a training package while listening to an oral explanation from the experimenters/instructors.

Students in the detailed training group had the above described materials plus an additional set of exercises for each of the six linkage groups and for the map drawing instruction. This additional practice required another 50-minute training session to complete.

After simultaneously finishing the training treatments (the detailed training group started a day earlier than the brief training group), both groups received the linkage test and the drawing test.

Scoring. The linkage scores measured the students' ability to identify the relationship between 14 pairs of concepts. Ten points were given to every correct proposition.

Scores on the drawing test indexed the students' ability to integrate the 14 propositions into a map. Ten points were assigned to each correctly integrated proposition. The maximum score for the linkage and drawing test measures was 140.

In preparation for analyzing the data in Phase II, the students were classified into one of two groups: those who had learned to map well (good-mappers) and those who did not learn to map well (poor-mappers). This was done by calculating a composite post-training mapping test score, using linkage scores and drawing scores. Raw scores from both measures were transformed to t-scores and summed to form the composite score. Then, the median composite score was used to divide the students into two groups. This categorizing process appeared to be very appropriate in that the distribution of post-training scores was rather U-shaped with only a few students clustered in the middle.

Research Design

A two group comparison was used to test the effects of the training procedures. The independent variable had two levels of training: brief and detailed training. The dependent measures were scores on the linkage test and the drawing test.

The Chi-square median test (Glass & Hopkins, 1984) was used to test the statistical significance of the difference between the median performance of the two independent groups. A non-parametric statistical test was used because the distribution of scores was extremely non-normal and severely violated the assumptions required to use parametric tests.

RESULTS

Phase I: Determining the Effects of Training on Learning to Map

The post-training linkage scores and drawing scores were used to determine the effectiveness of the mapping training. There were 68 students in the brief training group and 63 in the detailed training group. The results showed that the linkage scores were negatively skewed. In addition, at least half of the students could answer more than half of the questions correctly. Also, the negative skewness and U-shape distribution of the drawing scores illustrate the fact that both the high- and low- scores were clustered together, but there were more high scores than low scores. Because the distributions of these two measures did not meet the basic assumptions of parametric inferential statistics, two separate Chi-square median tests were used to test the significance of the training effects.

Results from the *linkage* scores showed that 41% of the briefly trained students and 59% of the detailed trained students scored above the common median. The Chi-square statistic from the ensuing median test equaled 4.67 and was significant at the 0.05 level. Results from the *drawing* scores showed that 36% of the briefly trained students and 64% of the detailed trained students scored above the common median. This Chi-square statistic from the ensuing median test equaled 11.62 and was significant at the 0.001 level.

Phase II: Determining the Effects of Reading, Instruction, Text Availability and Student Ability on Mapping Performance

METHOD

Subjects

Data from all 131 students described in Phase I were used in this Phase.

Materials

Text and Instructional Materials. The subject matter of the materials was based on a science topic that is common to almost all middle-school science curricula--the structure and functions of the leaf in green plants. The materials included: (a) a 500- word passage, (b) a 500-word script for oral instruction and (c) 21 color slides. For the no-instruction group, a filler passage titled, "Energy," was used.

The Post Instructional Mapping Test. Based on information from the instructional materials, a post-instructional mapping test was produced. This mapping test consisted of a title, 15 concepts, six linkages and a skeleton map. The test was titled, "The structure and functions of a leaf in green plants," used the following concepts: cell respiration, chlorophyll, chloroplasts, growth, leaf, lower epidermis, palisade layer, photosynthesis, plant, spongy layer, sugar, transpiration, water vapor loss and waxy cuticle and six linkages: has a (is a part of), is characteristic of, is an example of, results in, enables, is next to (takes place in). Figures 2 and 3 show maps which have been redrawn, but were originally constructed by two students under testing conditions in the study.

[Insert Figures 2 and 3 about here.]

The Post-Instructional Short-Answer Test. The Short-Answer test was composed of seven literal, two adjacent inferential, four non-adjacent inferential and two summarizing questions, after Pearson and Johnson (1978). Literal questions were those that required specific information in the text which was located in one sentence. Inferential adjacent questions required information from two adjacent sentences. Inferential non-adjacent questions required information from at least two sentences that were not adjacent. Summarizing questions required information from several parts of the passage.

Note that each of the passages and tests were developed with great care. To ensure this, three independent content area experts judged whether the passage and test items adequately and accurately sampled the domain of interest--the structure and functions of the leaf. After collecting and summarizing comments from the experts, the passage was revised three times, the last time by a curriculum and instruction expert. Similarly, test concepts for the mapping test and test items for the short answer test were also revised two and three times, respectively. The master (the map that was used as a test key) was developed by two other content area experts.

The Questionnaire of Attitudes Toward Taking A Mapping Test. In order to explore the students' reactions to this new type of testing, a questionnaire with nine items using Likert-type rating scales was used. See Table 4 for a listing of the nine items.

[Insert Table 1 about here.]

Procedures

All students were randomly assigned to one of the three instructional treatment groups (No-Instruction, Read Only and Read Plus Slides). Further, the students in the Read-Only and the Read Plus Slides groups were randomly divided into groups that took the posttest with or without accessibility to the text. All students in the No-Instruction group took the post test without having access to the text.

Students in the No-Instruction group were not given any information related to the structure and functions of the leaf, but were required to read a test-unrelated passage, and to take the post-instructional mapping test using only their prior knowledge of the content. Students in the Read Only group were instructed to read a 500-word passage related to the structure and functions of the leaf for 20 minutes. Students in the Read Plus Slides group listened to the oral instruction from the experimenters/instructors and viewed the corresponding slides for about 6 minutes. Then, they read the same 500-word passage as did the students in the Read-Only condition. Finally, for a filler task before taking the test, all students worked crossword puzzles for 10-minutes which was related to the passage that they had been reading.

After completing the instructional treatments, all students took the post-instructional mapping test and the short answer test. After completing both tests, the attitude questionnaire was administered.

Scoring Procedures. To initiate this process each student's map was rewritten into a list of propositions. The propositions were then classified into 20 different accuracy categories (see Table 2). These categories were based on three aspects of each proposition: (a) the pair of concepts, (b) the linkage that was used to label the relationship between the two concepts, and (c) the direction in which the arrow pointed.

[Insert Table 2 about here.]

These procedures were used to score each map. First, each *pair of concepts* provided by the students could fall into one of three accuracy levels. For the first level, the concept pair provided by the student was exactly the same as that provided by the experts on the "master map." For the second level, the concept pair differed from the standard map but was considered to be partially correct. For example, the master map showed that, **LEAF** has a **SPONGY LAYER**, but if a student showed that **PLANT** has a **SPONGY LAYER**, then it was considered to be partially correct. This latter proposition is true, but it does not represent the most informative way to represent the relationship between **PLANT** and **SPONGY LAYER**, for example **PLANT** has a **LEAF** has a **SPONGY LAYER**. A third level of accuracy was represented by a concept pair which differed from the master map and was considered to be incorrect, for example **WAXY CUTICLE** has a **SPONGY LAYER**.

Thus, when students included the same two concepts in a proposition as found on the master map, we referred to them as adjacent concept pairs. If the student formed a proposition with concept pairs other than those found on the master map, we referred to them as non-adjacent concept pairs. For example, **LEAF** and **SPONGY LAYER** is an adjacent concept pair, while **PLANT** and **SPONGY LAYER** is a non-adjacent concept pair.

Second, the *linkage* used by the student to label the relationship in a proposition could be correct, partially correct, wrong or missing. The partially correct label was limited to those propositions requiring the two cause-effect types of linkages, which have only subtle differences to eighth graders. For example, students might use "results in" instead of "enables" to show the relationship between a pair of concepts.

Third, the *direction* of the arrow drawn by students could be: (a) correct, (b) wrong, or (c) missing.

Table 1 shows the various possibilities of scoring combinations that were used in this study. A weighting scheme was designed that worked like this: If a student's proposition included the same concepts as found in a proposition on the master map (adjacent concepts) *and* the correct linkage, then 4 points were awarded for that proposition. If the linkage label was only partially correct, 2 points were awarded. If the linkage label was wrong or missing then no points were awarded. In addition to the above scheme, if the arrow between the two concepts pointed in the correct direction, then 2 points were awarded, but if the arrow was missing or pointed in the wrong direction, no points were added. So, a proposition formed with an adjacent concept pair could attain scores of 6, 4, 2 or zero points, depending on its degree of accuracy.

In addition, propositions formed with non-adjacent concept pairs had a similar weighted scoring system. If the linkage between the non-adjacent concepts was labeled correctly then 2 points were awarded. In addition, if the arrow pointed in the correct direction, 1 point was awarded. Thus, propositions formed with non-adjacent concept pairs could attain scores of 3, 2, 1, or zero points. Third, propositions which included concepts or linkages which were not listed on the test received zero points.

In addition to giving points to each proposition, we classifie each one according to a knowledge classification scheme. This scheme has four types: correct, partially correct, lack of knowledge and faulty knowledge. See Table 3 for explanations of these knowledge types.

[Insert Table 3 about here.]

The short-answer test. Based on answer keys developed by the experimenters and the content area experts, different weights were assigned to each question depending on how much information was needed for each question.

Questionnaire. The ratings chosen for each item were used as scores. That is, if a student indicated a "2" for item 3, then a "2" was used as a score.

Research Design

Theoretically, a 3x2x2x2 factorial design was used to estimate the effects of reading/instruction, availability of text while taking the test, mapping ability and two levels of student general ability on Post-Instruction test scores. However, it was pedagogically impossible to fill all the cells in the design (students in the No-Instruction group did not have text available to them when they took the exam) which resulted in an incomplete factorial design. Four of the cells were empty.

The *reader/instruction* factor had three levels: No-Instruction, Read-Only, and Read Plus Slides. The *text availability* factor had two levels: taking the test without access to the text vs. taking the test with text available. The student *general ability* and *mapping ability* factors had two levels each: high ability vs. low ability. The general ability of students was determined by dividing the students' scores on the Otis-Lennon Mental Ability Tests at the median, and the mapping ability levels were based on the students' performance on the training phase of this study.

Measures from the post-instruction mapping test and the post-instruction short answer test were used as dependent indices of learning in the experimental plan. Scores from the attitude questionnaire were used as a dependent measure in the experimental plan, as well. The two test scores were analyzed as repeated measures in the general linear model plan discussed by Wilkinson (1987). This repeated measures variable was called the *test type* factor and had two levels: the mapping test and the short answer test.

Since some of the independent variables were interdependent due to unequal cell sizes and to the partitioning of the general and mapping ability variables, the order of adding main effect variables to the equation followed this plan. First, the variance due to the students' general ability was removed, followed by the variance due to their mapping ability. Next, the effects of reading/instruction and of text availability were removed. Finally, the significant interaction terms were added in. This plan was designed to give the instruction/reading and text availability factors the most stringent test since they were added to the analysis after major chunks of variance due to the correlated ability factors (mapping and general abilities) had been removed.

RESULTS

Phase II: Determining the Effects of Reading, Instruction, Text Availability and Student Ability on Mapping Performance

Performance

The ANOVA results are portrayed in Figures 4 and 5. Please note that all main effect and interaction results refer to a repeated measures analysis unless indicated to the contrary. The results showed that the good mappers scored significantly higher than the poorer mappers ($F = 36.6$, $df = 1,124$, $p < .0001$), the students that had the text available when they took the test scored better than those that did not ($F = 42.4$, $df = 1,124$, $p < 0.0001$) and the higher ability students scored higher than the lower ability students ($F = 36.6$, $df = 1,124$, $p < .0001$). The main effect due to reading/instruction was not significant ($F = 1.88$, $df = 2,124$, $p < .17$) when considered in the repeated measures analysis. However, the univariate analysis showed a main effect for reading/instruction ($F = 4.7$, $df = 2,124$, $p < 0.01$) when using the mapping scores, but the effect was not significant ($F = .55$, $df = 2,124$, $p < .58$) when using the short answer scores.

[Insert Figures 4 and 5 about here.]

In addition, Bonferroni multiple comparison techniques (Wilkinson, 1987) indicated that students who either read the passage or saw the slide show scored higher than the group that received no instruction, ($F = 8.0$, $df = 2,123$, $p = .005$ and $F = 8.15$, df - $2,123$, $p = .005$, respectively for the mapping scores only). Multiple comparison analyses showed no difference in the performance of the two groups that received the two types of instruction.

The student ability by text interaction was significant ($F = 9.27$, $df = 1,124$, $p < .01$). Finally, the text availability by test type interaction (mapping vs. short answer scores) was significant ($F = 297.9$, $df = 1,124$, $p < .0001$).

The student ability by text availability interaction was due to the tendency of the higher ability students to use the text better when it was available during the test than were the lower ability students. The text availability by test type interaction was due to the fact that the students made better use of the text when it was available on the short answer test than they did when it was available on the mapping test. However, having the text available while taking the test was beneficial to students on the mapping test ($F = 4.1$, $df = 1,134$, $p < .04$) and on the short answer test ($F = 54.5$, $df = 1,124$, $p < .0001$), as well.

Correlation coefficients between pairs of the six school measures and the two post-instruction test measures are shown in Table 4. Complete data from 52 students who learned to map well were used to estimate the correlation coefficients. To estimate correlations, a coefficient between each of the school measures and the mapping and short answer test scores was computed for students in each cell in the experimental design (summing over the ability factor cells). These five correlation coefficients were averaged using a Fisher's transformation routine. Each pair was then tested to determined if there was a significant difference between the correlation with the mapping test and the correlation with the short answer test using a procedure described by Glass & Hopkins (1984).

[Insert Table 4 about here.]

An inspection of Table 4 shows that the correlation between the mapping test and each school measure is larger than the correlation between the short answer test and that same school measure. The magnitude of these differences is statistically significant for the Otis-Lennon Mental Ability Test ($p < .002$) and marginally significant for the Stanford Vocabulary Test ($p < 0.10$), while the remaining differences are not statistically significant.

Analysis of Propositions

The item difficulty level (percentage of students who got that item entirely correct, e.g., no part scores) of each of the 20 pairs of adjacent concepts (propositions) was computed. Further, an instructional sensitivity index was computed for each item by comparing the percentage of students' correct responses who had *no* instruction with those who read and/or saw the slides under two conditions: access to text during testing and without access to the text during testing. A positive sensitivity index shows that the item is sensitive to the increase in knowledge due to reading/instruction. A negative sensitivity index indicates that the reading/instructional event may have confused the student about the content measured by that test item (see Table 5).

[Insert Table 5 about here.]

The results showed that when the students did not have access to their texts during the test, there were 5 items (propositions) with a negative sensitivity index and 15 with a positive index. When the students had access to the text during the test, only 1 item had a negative sensitivity index and 19 had a positive index. The two highest sensitivity indices noted were .52 and .55.

Questionnaire Results

The students' responses on the post-instructional questionnaire are shown in Table 1. Results showed (Items 1, 3, and 4) that taking the mapping test was a new experience to the students. However, the poor-mappers, more than the good mappers, indicated that they had difficulty understanding the meaning of the linkages and/or following the instructions to construct a map.

According to the responses to items 2, 5, 6, 7, 8, and 9, the students reported that: (a) taking the mapping test did not seem to interest them, (b) they did not prefer to take a mapping test rather than answer short answer, multiple-choice or true-false questions, (c) they would not like for the mapping test to become a new type of test, and (d) they had neither positive nor negative opinions about how helpful the test was to see how these concepts were organized.

CONCLUSIONS

Criteria

We refer back to the criteria in the introduction and explain how compatible these results are with those criteria.

Criterion 1. The 'technology' of taking the test should be simple enough so that all students can learn to do it, even the slower and less-motivated ones.

The results related to this criterion come from three sources: (a) the characteristics of central tendency and dispersion of both the post-training mapping test measures (i.e., the linkage score and drawing score); (b) the relationships of these two scores with the Otis-Lennon School Ability Test and other achievement scores and school grades, and (c) the difference between the means of the two post-training mapping test scores. These sources indicate that many of the students learned to map.

The majority of students learned to draw a map but there were still many who did not. With the model drawing score of 140, which equals the maximum possible drawing score, the U-shaped distribution of drawing scores supports the notion that once the students learned how to draw a map, they tended to be able to do so very well. On the other hand, if the students had not learned to draw a map, they got a relatively low score which was close to zero.

The low to moderate correlation coefficients of the two post-training mapping test scores (the linkage scores and drawing scores) with the school ability and achievement test scores/grades for both the brief training group and the detailed training group suggest that the linkage score and drawing score measured somewhat different abilities from those measured by school tests and grades. The weak relationships between linkage scores and drawing scores for both the brief and the detailed training group suggest that mapping ability involves at least two separate subskills.

The range of the Otis-Lennon School Ability Test scores for good-mappers was from 85 to 141, whereas the same kind of test scores for poor-mappers ranged from 81 to 121. These data illustrate that both good- and poor-mappers came from students of widely dispersed abilities.

There was a clear training effect between the brief and the detailed training group as evidenced in the median test for both the linkage and the drawing scores. The major treatment difference between the brief and the detailed training group was that the former group did not receive the one day extra practice. The extra practice day increased the post-training mapping test scores by approximately 27%. These data add evidence to the assertion that most students can be taught to map, especially if they receive practice.

While the case seems clear that most of these students have the potential to learn to map, the fact remains that many did not. One possible explanation is supported by the fact that there was a significant mapping ability effect, although weak, when the *short answer* test scores were used as the dependent measure. There is no good reason why the short answer scores should show a mapping ability effect unless the index of mapping ability has other factors confounded within it. A very likely one probably relates to task motivation.

We suspect that the requirements to map a set of ideas like these are too difficult for some of these students. This task requires an analytic problem solving approach that is, perhaps, foreign to many of these students. Also, it may overemphasize a "bottom-up" strategy at the expense of a "top-down" strategy. In other words, the students may be spending more time putting propositions together than trying to construct the macrolevel relationships. We anticipate that when other models of knowledge structure continue to be used for assessment purposes, like those by Surber (1984) and Novak et al. (1983), then the formidable task of constructing a map will be replaced by tasks of selecting "frames" and filling in slots. These tasks appear to be less demanding than those associated with constructing a map.

Criterion 2. The technique must be sensitive to reading/instruction. That is, an increase in students' test scores should be concomitant with an increase in students' knowledge.

The results indicate that there was a reading/instructional effect as determined by the significant difference between the performance of students who received no instruction with the performance of those who read and/or saw the slides. However, there was no difference between the performance of the students who simply read the passage and those who saw the slides.

In comparison, there was not a reading/instruction effect on the post-instruction short answer test. This result suggests that the mapping test might be more sensitive to reading/instruction than a 'regular' classroom type test. At least, it can be said that the mapping test was no worse than the short answer test when they were used to measure the students' knowledge growth.

Criterion 3. Students should be able to score better on the test when a text containing information being tested is available at testing time.

The results indicated that there was a clear text availability effect, although it is complicated by some interactions. The significant text availability by test type interaction is due to the fact that having the text was very facilitative for the students when they took the short answer test, while less significantly so when they took the mapping test. In addition, there was a significant ability by text availability by test type triple interaction. As indicated in Figures 1 and 2, while the higher ability students were able to make good use of the text, irrespective of which test they were taking, lower ability students were able to use the text well for the short answer test but not for the mapping test. A possible explanation is that the text information had to be located and "translated" into propositional terminology when taking the mapping test, but could be used "directly" and often literally in the short answer test.

Another possible explanation is that the students took the short answer test after they had finished the mapping test. So, they read, and may have referred to the text many times by the time they answered the short answer questions with the text available.

Criterion 4. Performance on the test should be comparable to (correlate with) performance on other acceptable testing and grading techniques.

The most direct comparison between mapping scores and other test scores is with the short answer test. The mapping test appears to be more sensitive to reading/instruction than the short answer test, while the short answer test is more sensitive to the availability of text (at least with the higher ability students) than is the mapping test.

In addition, the correlation coefficients showing the linear relationship between the mapping test and other student ability and achievement scores/grades were no less than, and not significantly different from those of the short answer test. Further, the correlation coefficients are in the moderate range of linear relationships. These results imply that these two types of assessment techniques might measure different characteristics of the students' ability while they still measure many common ones. These results are evidence that the mapping test is concurrently valid enough, according to a variety of criteria, that teachers could justify its use.

Criterion 5. The procedures and products of an assessment technique should be consistent with the instructional goals and practices of the school, and the integrity of the content area.

This criterion will be discussed under four topics: content validity, item validity, usability of the tests and diagnostic value.

Content validity. Content validity is an extremely important but difficult test characteristic to demonstrate. In this case we defined content validity as the appropriateness of the content area, the passage, and the test items. To help ensure content validity, several steps were taken. First, the passage was chosen from junior high school life science textbooks, and the tests were derived from information in that passage.

Second, a panel of three independent experts as mentioned earlier (other than the items writers) judged whether the passage and test items adequately and accurately sampled the domain of interest--the structure and functions of the leaf. Also, the standard map was developed by two other content area experts.

Item Validity. There are several ways to determine the validity of individual test items. One of these possibilities is called the measure of instructional sensitivity. The measure of instructional sensitivity is "a measure of how well an item discriminates between examinees who have received instruction and those who have not" (Crocker & Algina, 1986, p. 330).

The results showed that the content material of the mapping test was relatively difficult for the students, even for those who read the passage and/or saw the slide presentation. This is evidenced by the fact that only 4 out of 20 propositions had difficulty levels which were higher than 0.50. These results were consistent with the low scores on the short answer test, that is, an average score of 46.9 with the maximum possible score of 200.

There are several possible reasons to explain this low performance. First, the reading/instruction time was relatively short. Second, some students lacked motivation to learn since it was obvious to them that this whole process was just a university conducted research activity and they had a right to choose not to respond very enthusiastically. Third, although the content of the passage and slides was selected from junior high level textbooks, the abstract level of the content may have contributed to the difficulty level. For example, data from Table 4, showed that the propositions which were related to the *structure* of the leaf (a rather concrete set of ideas) were easier than those propositions related to the *functions* of the leaf (a rather abstract set of ideas). Propositions 1 to 10 involved the structure of the leaf, whereas Propositions 11 to 20 were related to the functions of the leaf.

Usability of the tests. The test administrators reported that they did not have difficulty administering the mapping test. The time required for training students how to take the mapping test did not seem exhaustive, but the students would probably have learned to map better if they were given more practice and/or small group instruction. Moreover, taking the mapping test itself did not require more time than any ordinary classroom test.

Although the mapping scoring scheme was rather objective, the complexity of the process might hinder a teachers' motivation to use it since transforming students' maps into a set of propositions takes time. So, instead of transforming the maps into propositions, a checklist with a list of adjacent nodes and their relative partially correct non-adjacent nodes could be developed to save teachers' time.

Also, we suspect that a "cloze" version of the test would be significantly easier to grade. This means that selected concepts and linkages from the standard (key) map could be deleted and the students' job on the test would be to reconstruct those deleted items. The teacher's grading task would be considerably simplified with the use of these "cloze" maps. Whether or not these "cloze" maps are as sensitive to reading/instruction as the constructed kinds remains a research question.

Based on the students' responses on the questionnaire, the good-mappers did not have difficulty identifying the linkages and drawing a map, whereas the poor-mappers responded that they had difficulty doing so. However, the results of the post-training test showed that most students could be taught to map, especially if they had the opportunity to practice.

According to the responses on the questionnaire, the students had no strong opinion about whether the mapping test had a positive influence on their learning. Unfortunately, they did not seem to like this kind of assessment instrument. Since there were no questions in the questionnaire that gave the students an opportunity to explain why they did not like it, this issue remains unclear. We suspect, however, that the technique was too unfamiliar and/or too much work under the circumstances for most students' preference. We suspect, however, that the students would have a more positive attitude toward taking a "cloze" map test.

Diagnostic value. Another important aspect to consider when using an assessment technique is its potential diagnostic value. Based on the scoring schemes used in this study and the maps produced by the students, a relatively detailed and holistic profile of each student's performance was obtained.

For example, Figure 2 is a map drawn by a student who received no instruction. Table 6 was derived from Figure 3 and shows (a) a set of propositions and (b) the student's accuracy and deviation score based on the scoring schemes used in this study. Further, the set of propositions is divided into four subsets of propositions with respect to correct knowledge, partial knowledge, faulty knowledge, and lack of knowledge.

[Insert Table 6 about here.]

From Table 6, the following diagnosis can be made about this student's knowledge. First, this student knew that the LEAF is a part of the PLANT and that both the PALISADE LAYER and the LOWER EPIDERMIS are parts of the LEAF, whereas she did not know that the WAXY CUTICLE and the SPONGY LAYER are also parts of the LEAF. Instead, her map showed that the WAXY CUTICLE and the SPONGY LAYER are parts of the PLANT, which is less specific and accurate information than those corresponding propositions developed by experts. Similarly, another proposition, CHLOROPHYLL is a part of LEAF, exposed the same kind of problem. This student seemed to know that there was a relationship between CHLOROPLASTS and CHLOROPHYLL, but put the wrong linkage between them. Third, this student did not know that CHLOROPLASTS can be found in the PALISADE LAYER and the SPONGY LAYER. Fourth, this student was not aware of the spatial relationships among the layers of the leaf.

This information suggests that the student knew parts of the big picture about the structure of the leaf (i.e., PLANT, LEAF, PALISADE LAYER, and SPONGY LAYER), but had little knowledge about the smaller parts of the leaf (i.e., CHLOROPLASTS and CHLOROPHYLL), and could not identify the location of the parts of the leaf.

Information about the functions of the plant suggests that she had the following knowledge problems: (a) Even though she correctly stated that GROWTH takes place in PLANT, she did not know that SUGAR enables GROWTH and CELL RESPIRATION can result in GROWTH. (b) The student's proposition, CELL RESPIRATION takes place in PALISADE LAYER indicated that she did not know that CELL RESPIRATION takes place in every cell of the PLANT. So, she only had partial knowledge about the function and location of cell respiration. (c) The student mapped that PHOTOSYNTHESIS takes place in PLANT which was less accurate than the proposition PHOTOSYNTHESIS takes place in CHLOROPLASTS. Moreover, this student did not seem to know that CHLOROPHYLL enables PHOTOSYNTHESIS. (d) She did not have knowledge about the function of transpiration since neither of these propositions--TRANSPIRATION takes place in LOWER EPIDERMIS and TRANSPIRATION results in WATER VAPOR LOSS--was produced by her. Finally, it showed that this student had faulty knowledge about the relationship between the concepts of SUGAR and CHLOROPHYLL.

The next example is from a student who read the passage, saw the slides and took the test with the text available. This student's map and its related information are shown in Figure 3 and Table 7. Figure 3 and Table 7 were constructed using the same procedures and rules as those used in Figure 2 and Table 6, respectively.

[Insert Table 7 about here.]

From Figure 3 and Table 7, it can be seen that this student mapped 14 correct out of 20 possible propositions. She seemed to have a rather complete knowledge about the structure of the leaf. An analysis of the leaf functions showed: (a) She had correct knowledge about the function of transpiration. (b) She only missed one proposition about photosynthesis, CHLOROPHYLL enables PHOTOSYNTHESIS. (c) She did not identify GROWTH as a characteristic of the PLANT, that CELL RESPIRATION takes place in PLANT, and that SUGAR enables GROWTH. In addition, she only had partial knowledge about the proposition that CELL RESPIRATION results in GROWTH. While she was rather clear about photosynthesis, she was still confused about the processes of plant growth and cell respiration.

Implications

The purpose of this study was to develop and evaluate a psychological instrument that was sensitive to the changes in content knowledge structure as a student reads text and/or receives instruction. Since the findings of this study are generally consistent with the purpose, four educational implications are proposed.

The first has to do with assessment. The mapping technique has the potential to become another approach to the classroom teacher's arsenal of instruments. Although some students seemingly have difficulty learning how to map, the majority of them can learn. We suspect, however, that were the teacher to embed the mapping procedure into the regular classroom teaching and learning program, the students' ability to map would grow consistently. Also, the use of "cloze" maps would also aid the learning process.

An additional implication is that a computerized version of mapping may help overcome some of the problems discussed earlier. Computerized mapping tests have certain potential advantages over the paper-and-pencil version such as: (a) students can draw a map more freely and efficiently because they can move the concepts, linkages, and arrows around and the computer will keep track of the patterns, (b) the scoring will be facilitated and human scoring errors can be minimized, (c) the profiles of the test results can be illustrated more comprehensively, and further diagnostic and remedial procedures will be easier to follow.

Also, the mapping technique seems to provide a visual, and easy way for students to understand the test results. When the standard map is shown to them and they are allowed to examine the differences between their map with the standard map, enlightened discussion will probably follow.

We offer some instructional implications also. We suspect that students' maps, item difficulty levels, and instructional sensitivity indices can be used to locate which segments of knowledge the students: (a) learned, (b) partially learned, (c) learned incorrectly, or (d) did not learn at all. Further, the teachers can use this information to speculate on reasons why this happened. Several possible reasons are: (a) the text did not provide necessary and sufficient information or provided incorrect information, (b) the teachers did not teach it or perhaps taught it incorrectly, and (c) the information was too complicated. For example, only a few students mapped the proposition that the "Waxy Cuticle is a part of Leaf." A close inspection of the text shows that this proposition was not stated very clearly. Similarly, CELL RESPIRATION results in GROWTH was very difficult to derive from the passage. An important educational implication of these observations is that parts of the text and/or the instruction need to be revised.

Implications for student learning include a reference to prior research by Holley and Dansereau (1984), and Armbruster and Anderson (1984) in which the process of mapping was demonstrated to positively affect the learning outcomes of reading and studying activities. We speculate that students can improve their understanding of the content by examining their mapping errors and correcting them by referring back to an informational source, like a textbook, or to a standard map. Also, we envision a computerized mapping test in which the program routine routes students into relevant instructional material so that they can correct their mapping errors. In this sense, the mapping technique can become a learning and/or note taking strategy as well as an assessment technique.

References

Armbruster, B. B., & Anderson, T. H. (1984). Mapping representing information text diagrammatically. In C. D. Holley, & D. F. Dansereau (Eds.), *Spatial learning strategies: Techniques, applications, and related issues* (pp. 189-209). New York: Academic Press.

Brachman, R. J. (1977). What's in a concept: Structural foundations for semantic networks. *International Journal of Man-Machine Studies, 9*, 127-152.

Champagne, A. B., Klopfer, L. E. DeSena, A. T., & Squires, D. A. (1978). *Content structure in science instructional materials and knowledge structure in students' memories* (Report No. 22). Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Champagne, A. B., Klopfer, L. E. DeSena, A. T., & Squires, D. A. (1981). Structural representations of students' knowledge before and after science instruction. *Journal of Research in Science Technology, 8*, 97-111.

Cohen, G. (1983). *The psychology of cognition* (2nd ed.). New York: Academic Press.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-247.

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology, 12*, 539-549.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

Dansereau, D. F., Collins, K. W., McDonald, B. A., Holley, C. D., Garland, J. C., Diekhoff, G., & Evans, S. H. (1979). Development and evaluation of a learning strategy training program. *Journal of Experimental Psychology, 71*, 64-73.

Geeslin, W. E., & Shavelson, R. J. (1975). An exploratory analysis of the representation of a mathematical structure in students' cognitive structures. *American Education Research Journal, 12*, 21-39.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Holley, C. D. (1979). *An evaluation of intact and embedded headings as schema cuing devices with non-narrative text*. Dissertation Abstracts International, 40 4491 A. (University Microfilms No. 80-02, 220)

Holley, C. D., & Dansereau, D. F. (1984). Networking: The technique and the empirical evidence. In C. D. Holley, & D. F. Dansereau (Eds.), *Spatial learning strategies: Techniques, applications, and related issues* (pp. 810-108). New York: Academic Press.

Klatzky, R. L. (1980). *Human memory: Structures and processes* (2nd ed.). San Francisco, CA: W. H. Freeman and Company.

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y-G., & Tucker, D. G. (1986). Inferring students' cognitive structures and their development using the "Ordered Tree Technique." *Journal of Educational Psychology, 78*, 130-140.

Novak, J. D., Gowin, D. B., & Johansen, G. H. (1983). The use of concept mapping and knowledge vee mapping with junior high school science students. *Science Education, 67*, 625-645.

Otis, A. S. & Lennon, R. T. (1977-79). *Otis-Lennon School Ability Test.* The Psychological Corporation, New York, New York.

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

Quillian, R. (1966). *Semantic memory.* Cambridge, MA: Bolt, Beranek and Newman.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 216-270). Cambridge, MA: MIT Press.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. (1972). A process model for long-term memory. In E. Tulving, & W. Donaldson (Eds.). *Organization of memory* (pp. 198-246). New York: Academic Press.

Rumelhart, D. E., & Norman, D. A. (1985). Representation of knowledge. In A. M. Aitkenhead, & J. M. Slack (Eds.), *Issues in cognitive modeling* (pp. 15-62). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Surber, J. R. (1984). Mapping as a testing and diagnostic device. In C. D. Holley, & D. F. Dansereau (Eds.), *Spatial learning strategies: Techniques, applications, and related issues* (pp. 213-233). New York: Academic Press.

Wilkinson, L. (1987). *SYSTAT: The system for statistics.* Evanston, IL: SYSTAT, Inc.

Table 1

Mean Scores from the Post-Instructional Questionnaire

Question	Poor Mappers M	Poor Mappers SD	Good Mappers M	Good Mappers SD
1. The mapping test was a new experience for me.	3.96	1.39	4.15	1.02
2. Taking the mapping test was interesting.	2.11	1.19	2.43	1.12
3. Had no difficulty understanding linkages.	2.42	1.32	2.95	1.41
4. Had no difficulty understanding instructions on how to draw a map.	2.56	1.29	3.26	1.42
5. Was helpful to see how concepts are organized.	2.68	1.32	2.94	1.22
6. Would rather take mapping test than short answer test.	2.59	1.49	2.62	1.43
7. Would rather take mapping test than multiple-choice test.	1.68	0.98	1.79	1.07
8. Would rather take mapping test than true-false test.	1.61	0.89	1.92	1.18
9. Would like for mapping test to be used at my school.	1.79	1.18	1.92	1.14

Table 2

The 20 Categories for Scoring Any Pair of Nodes

I. A Pair of Adjacent Nodes (Relative to the Standard Map)

	Linkage	Weight	Arrow	Weight	Total Score per Pair of Nodes
1.	Correct	4	Correct	2	6
2.	Correct	4	Wrong	0	4
3.	Correct	4	Missing	0	4
4.	Partially Correct	2	Correct	2	4
5.	Wrong	0	Correct	2	2
6.	Missing	0	Correct	2	2
7.	Wrong	0	Wrong	0	0
8.	Missing	0	Wrong	0	0
9.	Wrong	0	Missing	0	0
10.	Missing	0	Missing	0	0

II. A Pair of Non-Adjacent Nodes (Relative to the Standard Map)

	Linkage	Weight	Arrow	Weight	Total Score per Pair of Nodes
11.	Correct	2	Correct	1	3
12.	Correct	2	Wrong	0	2
13.	Correct	2	Missing	0	2
14.	Wrong	0	Correct	1	1
15.	Missing	0	Correct	1	1
16.	Wrong	0	Wrong	0	0
17.	Missing	0	Wrong	0	0
18.	Wrong	0	Missing	0	0

III. Others (Pair of Nodes is not relative to the Standard Map)

	Linkage	Weight	Arrow	Weight	Total Score per Pair of Nodes
19.	Wrong Pair of Nodes	0		0	0
20.	Adding Extra Concepts	0		0	0

Table 3

Classification of Four Types of Knowledge

Types of Knowledge	When Concepts Are: Adjacent		Non-adjacent	
	Linkage	Arrow	Linkage	Arrow
1. Correct	Correct	Correct	N/A	N/A
2. Partial	Partially Correct	Correct	Correct	Correct
	Correct	Wrong	Correct	Wrong
	Wrong	Correct	Wrong	Correct
3. Lack of	Missing	Missing	N/A	N/A
4. Faulty	Wrong	Wrong	Wrong	Wrong

Table 4

Pearson Correlation Coefficients Among Six School Measures, the Mapping and the Short Answer Test Measures

Measures	Mapping Test	Short Answer Test
1. Otis-Lennon School Ability	0.74	0.48
2. Stanford Vocab Achievement	0.64	0.48
3. Stanford Reading Comprehension	0.55	0.41
4. Stanford Science Achievement	0.66	0.64
5. Reading Grade in School	0.51	0.47
6. Science Grade in School	0.49	0.40
7. Mapping Test	----	0.69

Table 5

Item Difficulty Level and Instructional Sensitivity Index of Twenty Pairs of Adjacent Nodes on No-Instruction vs. Instruction Groups for Good Mappers

Types of Groups	No-Instruction during Test	Instruction without Text during Test		Instruction with Text	
No. of Cases	13	28		24	
	% of Students Pass	% of Students Pass	Sensitivity Index	% of Students Pass	Sensitivity Index
Adjacent Node Propositions					
1	.62	.64	.02	.71	.09
2	.15	.36	.21	.42	.27
3	.15	.46	.31	.67	.52
4	.38	.57	.19	.67	.29
5	.08	.43	.35	.63	.55
6	0	.07	.07	.25	.25
7	0	.04	.04	.38	.38
8	.15	.11	-.04	.25	.10
9	0	.07	.07	.17	.17
10	.08	.04	-.04	.42	.34
11	.08	.14	.06	.42	.34
12	.15	.24	.09	.58	.43
13	.15	.04	-.11	.21	.05
14	0	.07	.07	.13	.13
15	.23	.21	-.02	.13	-.10
16	.08	.18	.10	.33	.25
17	0	0	0	.17	.17
18	.23	.04	-.19	.38	.15
19	.08	.14	.06	.13	.05
20	0	0	0	0	0

Note. Sensitivity Index = % of Students from Instruction Groups Pass minus % of Students from No-Instruction Groups Pass.

Table 6

A Set of Propositions and Accuracy Scores Produced by a Student Who Received No Instruction and Took the Test Without Available Text

	Proposition Produced	Accuracy Score
1.	Leaf --- is a part of ---> Plant	6
2.	Palisade Layer --- is a part of ---> Leaf	6
3.	Lower Epidermis --- is a part of ---> Leaf	6
4.	Plant --- has a ---> Spongy Layer	3
5.	Plant --- has a ---> Waxy Cuticle	3
6.	Chlorophyll --- is a part of ---> Leaf	3
7.	Chloroplasts --- result in ---> Chlorophyll	2
8.	Growth --- takes place in ---> Plant	6
9.	Photosynthesis --- takes place in ---> Plant	3
10.	Cell Respiration --- is a part of ---> Palisade Layer	1
11.	Water Vapor Loss --- enables ---> Transpiration	0
12.	Sugar --- is characteristic of ---> Chlorophyll	0
13.	9 pairs of adjacent nodes were not produced by student A	0
Total Score		39

Table 7

A Set of Propositions and Accuracy Scores Produced by a Student Who Received Instruction and Took Test With Available Text

	Proposition Produced	Accuracy Score
1.	Leaf ---is a part of---> Plant	6
2.	Waxy Cuticle ---is a part of---> Leaf	6
3.	Spongy Layer ---is a part of---> Leaf	6
4.	Palisade Layer ---is a part of---> Leaf	6
5.	Lower Epidermis ---is a part of---> Leaf	6
6.	Palisade Layer ---is next to---> Layer	6
7.	Lower Epidermis ---is next to---> Spongy Layer	6
8.	Spongy Layer ---has---> Chloroplasts	6
9.	Palisade Layer ---has---> Chloroplasts	6
10.	Chloroplasts ---have---> Chlorophyll	6
11.	Photosynthesis ---takes place in---> Chloroplasts	6
12.	Photosynthesis ---results in---> Sugar	6
13.	Sugar ---is a part of---> Cell Respiration	2
14.	Cell Respiration ---enables---> Growth	4

Table 7 (Continued)

Proposition Produced	Accuracy Score
15. Water Vapor Loss ---takes place in---> Lower Epidermis	6
16. Transpiration ---results in---> Water Vapor Loss	6
17. 4 pairs of adjacent nodes were not produced by student B	0
Total Score	90

Figure Captions

Figure 1. Explanation and Examples of a Linkage.

Figure 2. A Map by a Student Who Received No Instruction and Took Test Without Available Text.

Figure 3. A Map Drawn by a Student Who Received Intensive Instruction and Took the Test with Available Text.

Figure 4. Mean Mapping Test Scores for Five Instructional Conditions and Two Ability Groups.

Figure 5. Mean Short Answer Test Scores for Five Instructional Conditions and Two Ability Groups.

LINKAGE: ATTRIBUTIVE RELATIONSHIP

... IS/ARE CHARACTERISTIC OF ...

<u>Meaning</u>: Y has a distinctive feature A (or X is characteristics of Y.) This feature X is used to describe, define, or explain Y.

Examples:

1. The color of human blood is red.

       is characteristic of
Red --------------------> Blood

2. Birds are the only animals that have feathers.

                    is characteristic of
Having feathers --------------------> Birds

3. Most snakes move by wriggling on their bellies.

                    is characteristic of
Belly wriggling --------------------> snake movement.

Figure 1

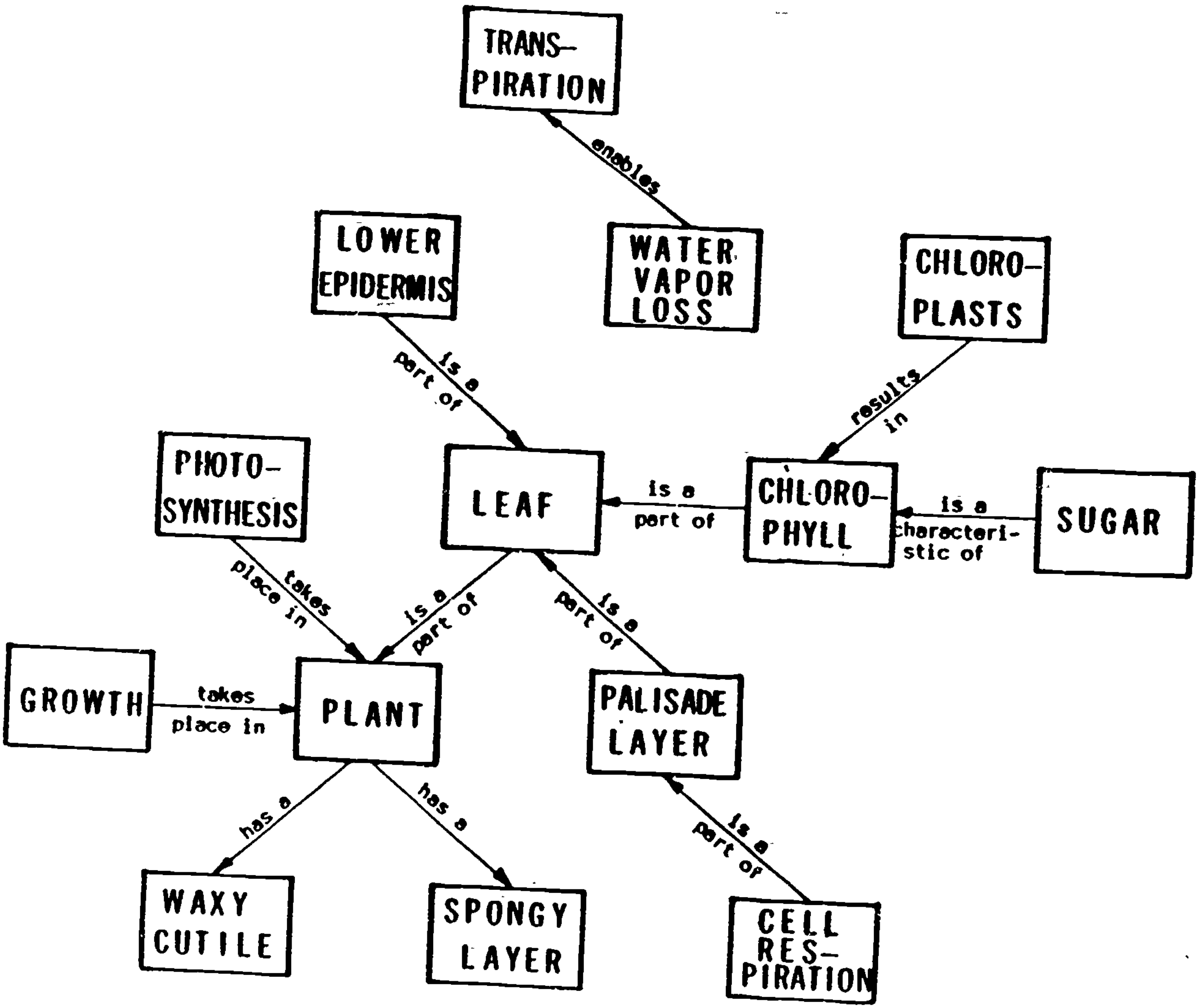

TRANS-
PIRATION
enables
WATER VAPOR LOSS
LOWER EPIDERMIS
is a part of
CHLORO-PLASTS
results in
PHOTO-SYNTHESIS
LEAF
is a part of
CHLORO-PHYLL
is a characteristic of
SUGAR
takes place in
is a part of
is a part of
GROWTH
takes place in
PLANT
PALISADE LAYER
has a
has a
is a part of
WAXY CUTILE
SPONGY LAYER
CELL RES-PIRATION

Figure 2

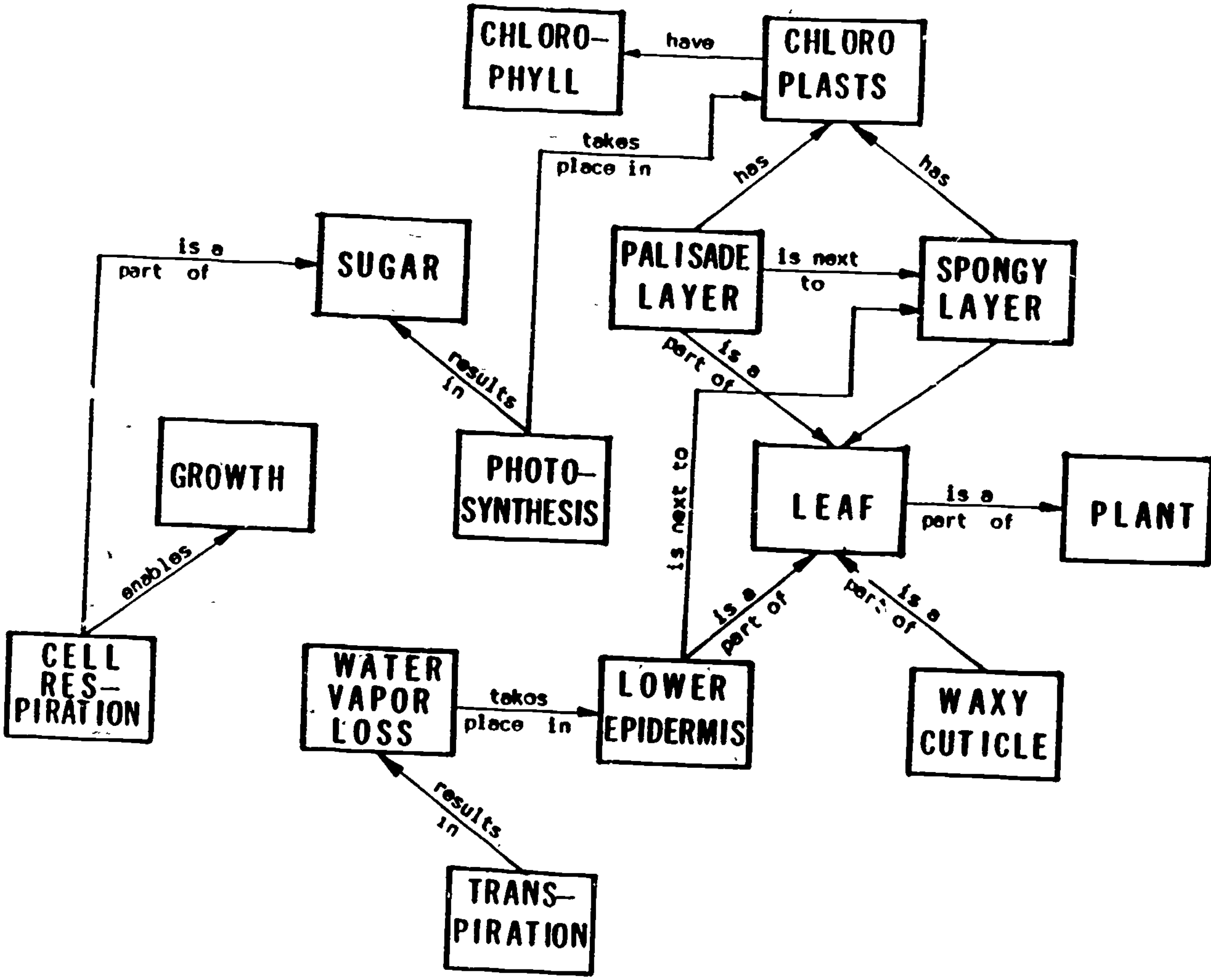

CHLORO-PHYLL
have
CHLORO PLASTS
takes place in
has
has
SUGAR
is a part of
PALISADE LAYER
is next to
SPONGY LAYER
results in
is a part of
GROWTH
PHOTO-SYNTHESIS
is next to
LEAF
is a part of
PLANT
enables
is a part of
is a part of
CELL RES-PIRATION
WATER VAPOR LOSS
takes place in
LOWER EPIDERMIS
WAXY CUTICLE
results in
TRANS-PIRATION

Figure 3

Figure 4

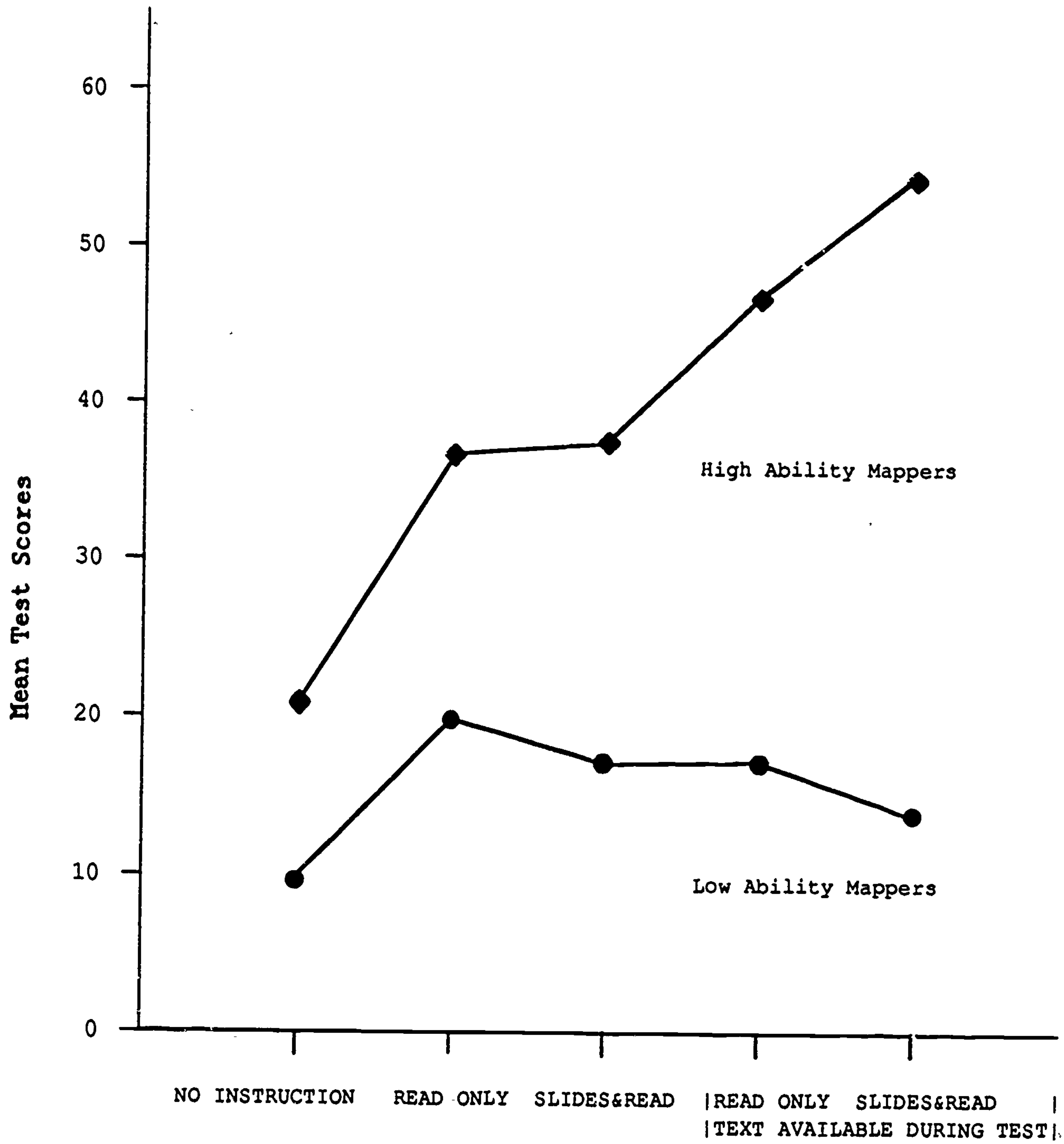
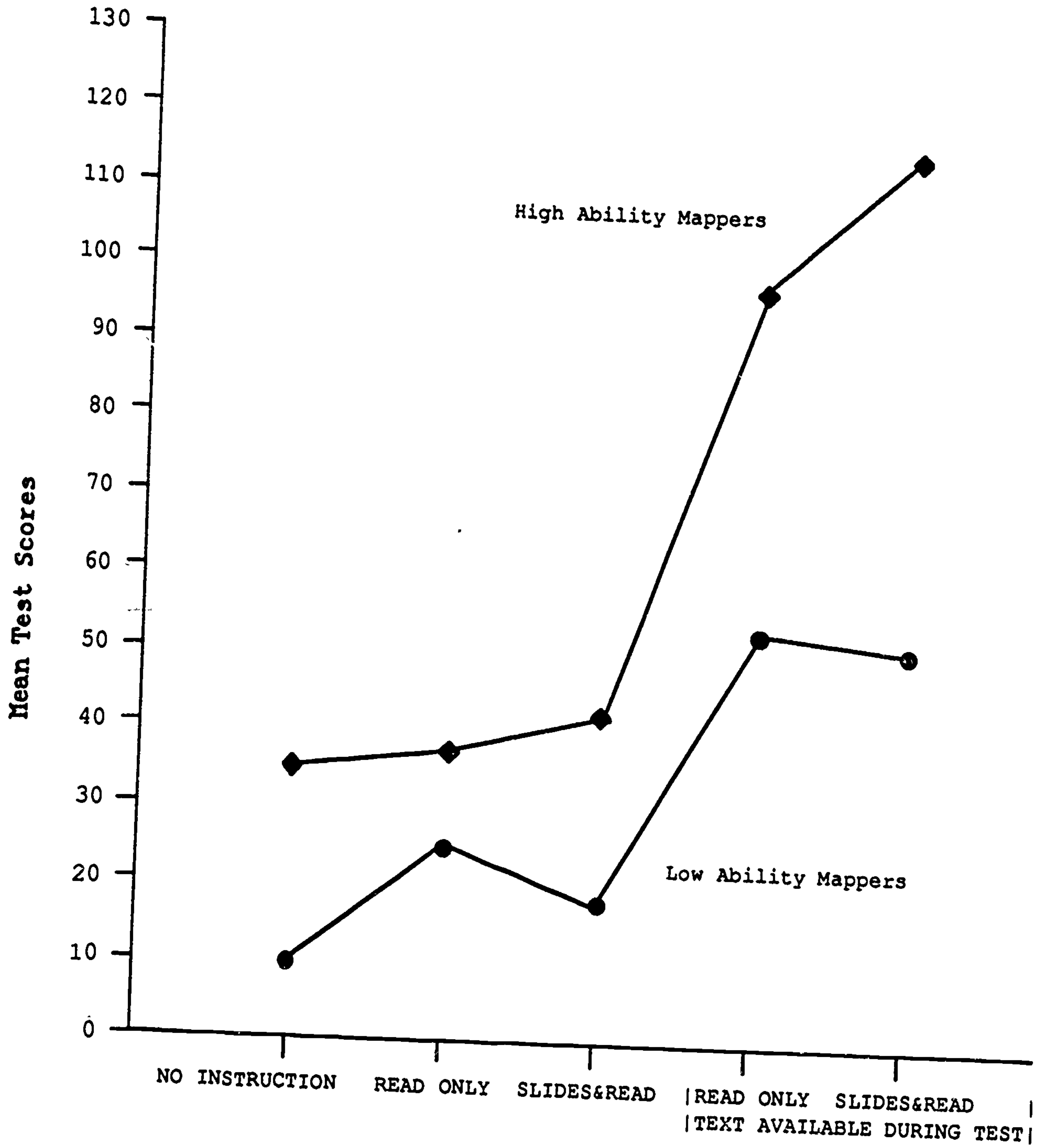

Mean Test Scores
60
50
40
30
20
10
0
High Ability Mappers
Low Ability Mappers
NO INSTRUCTION
READ ONLY
SLIDES&READ
READ ONLY
SLIDES&READ
TEXT AVAILABLE DURING TEST

Figure 5


130
120
110
100
90
80
70
60
50
40
30
20
10
0
Mean Test Scores
High Ability Mappers
Low Ability Mappers
NO INSTRUCTION
READ ONLY
SLIDES&READ
READ ONLY
SLIDES&READ
TEXT AVAILABLE DURING TEST