DOCUMENT RESUME

ED 408 200 SE 060 340

AUTHOR	Schau, Candace; And Others
TITLE	Use of Fill-in Concept Maps To Assess Middle School Students' Connected Understanding of Science.
INSTITUTION	Los Alamos National Lab., NM.
SPONS AGENCY	Department of Energy, Washington, D.C.
PUB DATE	97
NOTE	33p.; Paper presented at the Annual Meeting of the American Education Research Association (Chicago, IL, March 1997).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Cognitive Processes; *Concept Mapping; Junior High Schools; Middle Schools; *Scientific Concepts
IDENTIFIERS	*Middle School Students

ABSTRACT

This paper is based on the belief that knowledge must be organized in order to be accessible from long term memory and this kind of organization requires connected understanding. It reports on the development and validity of the select-and-fill-in concept map format designed to measure middle school students' connected understanding of selected science concepts and processes. The validity concerns discussed are related to map assessment content, cognitive processes required to complete the map assessment tasks, technical quality, and relationships of map scores and patterns of map scores to those from other assessment measures. Phase I included development and field testing of a variety of possible mapping formats. Phase II involved the creation, field testing, and revision of the most promising, the select-and-fill-in concept map, into its final form. In Phase III the format was tested with ethnically diverse middle school science students. Findings indicate that the select-and-fill-in concept map format can be used with ethnically diverse middle school students to measure their connected understanding of science. Contains 23 references. (Author/JRH)

Use of Fill-in Concept Maps to Assess Middle School Students' Connected Understanding of Science

Candace Schau, Nancy Mattern, & Robert W. Weber,
University of New Mexico;
Kirk Minnick, Minnick & Associates, Inc.
Connie Witt, Los Alamos National Laboratory

This paper was presented at the 1997 annual meeting of the American Educational Research Association, Chicago, IL. For additional copies of the paper or comments about it, contact Candace Schau, Professor; Psychological Foundations Program; College of Education; University of New Mexico; Albuquerque, NM 87131; 505-277-3208; cschau@unm.edu. This work was performed under contract to Los Alamos National Laboratory with funding provided by the U.S. Department of Energy. We would like to thank the teachers and their students who formed a critical part of our team.

Abstract

We believe that knowledge must be organized in order to be accessible from long term memory; this kind of organization requires connected understanding. If connected understanding is an important educational goal, we must be able to assess this kind of achievement. This article presents research in which we developed and began to explore the validity of the select-and-fill-in concept map format designed to measure middle school students' connected understanding of selected science concepts and processes. Phase I of our work included the development and field testing of a variety of possible mapping formats. Phase II involved the creation, field-testing, and revision of the most promising format, the select-and-fill-in concept map, into its final form. In Phase III, we tested this format with ethnically diverse middle school science students. In our work, we began to address validity concerns related to (a) map assessment content, (b) cognitive processes required to complete the map assessment tasks, (c) technical quality, and (d) relationships of map scores and patterns of map scores to those from other assessment measures. Our results indicated that we have taken some first steps toward showing that the select-and-fill-in concept map format can be used with ethnically diverse middle school students to measure their connected understanding of science.

Use of Fill-in Concept Maps to Assess Middle School
Students' Connected Understanding of Science

We believe that knowledge must be organized into mental networks into order to be accessible to the learner from long term memory. The current work on national standards and goals in the U.S. supports our belief in the importance of conceptual connections or connected understanding in science learning. The Benchmarks for Scientific Literacy (American Association for the Advancement of Science, 1993) explicitly emphasizes the importance of "coherence and connectedness" (p.XVI) in science learning, stating that "a central Project 2061 premise is that the useful knowledge people possess is richly interconnected" (p. 315). Similarly, the National Science Education Standards (National Research Council) indicates that "assessment processes that include all outcomes for student achievement must probe the extent and organization of a student's knowledge" (p. 82).

The concept of a schema serves as a useful component in many models of mental networks. Schemas are mental storage mechanisms that are structured as networks of knowledge (Marshall, 1995). One critical defining feature of a schema is the presence of connections; in order for a schema to be useful, the components within a schema must be interconnected. Schemas and connected groups of schemas often are called mental networks or cognitive structures.

Applying Marshall's (1995) concepts of schemas and mental networks to science learning suggests that expertise is attained by developing a mental network that contains rich, accurate, and relevant sets of interconnected science schema and schema components. Students learn science through 1) connecting new information about science into their exiting cognitive networks, 2) forming new connections among scientific information that already exists in their networks, 3) reorganizing their connected science schema to match incoming information, and 4) eliminating incorrect science concepts and connections.

Models of connected understanding (traditionally called structural knowledge) can be represented explicitly by visual-spatial networks or maps (see Figure 1 for an example). These maps represent one of the creator's mental models made explicit. In these maps, the concepts often are referred to as nodes and placed in a geometric shape, usually an oval or rectangle. Their connecting structures are called links and usually are represented by lines or arrows. A proposition consists of two connected science concepts and is the basic unit in structural knowledge (see Jonassen, Beissner, & Yacci, 1993, and Shavelson, Lang, & Lewin, 1994). For example, in Figure 1, the proposition "Earth revolves around sun" consists of the concepts of "Earth" and "sun" connected by the link "revolves around." Clusters of propositions that are more closely related to each other than to other propositions are connected into "neighborhoods" in these networks.

Insert Figure 1 about here

Assessing Connected Understanding

If connected understanding is an important educational goal, researchers and teachers must be able to assess this kind of achievement. As Ruiz-Primo and Shavelson (1996) noted, two general sets of techniques have been used to assess connected understanding. One set includes the "traditional" structural knowledge assessment research techniques; the second set is based on visual-spatial maps.

Traditional Assessments

The traditional research assessments were developed and continue to be used primarily by researchers interested in cognitive structures, often experimental psychologists. The general assessment process used involves first eliciting students' structural knowledge indirectly, often through the use of word associations or ratings of the similarity or relatedness of pairs of concepts from the field of interest. The student's structural knowledge can be characterized both visually (in the form of a map with unlabeled links) and numerically (as a score indicating the adequacy of the map, often in comparison to responses from experts). Both characterizations are obtained through computer software programs, such as Pathfinder. See Naveh-Benjamin, McKeachie, Lin and Tucker (1986) and Naveh-Benjamin, McKeachie, and Lin (1989) for an approach based on associations and Goldsmith, Johnson, and Acton (1991) and Johnson, Goldsmith, and Teague (1995) for an approach based on concept-relatedness ratings.

There is increasingly evidence of the validity of this class of approaches, although the evidence certainly is not complete. See, for example, Acton, Johnson, and Goldsmith (1994); Goldsmith and Johnson (1990); and Naveh-Benjamin et al. (1986).

However, these techniques have a number of limitations for use in the context of a classroom. First, they require computer software for analysis and so neither teachers nor students immediately see the maps that represent the students' connected understanding. Second, it is difficult for teachers and students to see how these techniques could assess something as important as structural knowledge, thus motivation to complete these formats can be low. Third, when students' structural knowledge is presented in a map, the links are unlabeled; labeled links are an important aspect of teaching and learning for connected understanding. Fourth, most (if not all) published research describing these approaches has been done with college students. At this point in their development, these kinds of techniques make better experimental research tools than classroom assessment measures.

Visual-spatial Map Assessments

The second general set of techniques used to assess students' connected understanding is based on visual-spatial maps, most often concept maps. As Ruiz-Primo and Shavelson (1996) indicated, a wide variety of task characteristics are associated with the use of concept maps as an assessment approach. Even so, the task demands of map assessments fall into three main categories. In the first, students are asked to generate their own maps. In the second, students complete partial maps through a fill-in process. In the third, the maps are created by someone else from student essays or interviews. In all three approaches, the resulting maps then are scored. For classroom purposes, the last approach is not very feasible.

Map Generation

In the most widely used assessment application, students generate their own maps. There is a great variety of approaches associated with this kind of assessment. For example, students may either generate their own concepts or use concepts that are given to them. This work also can be done individually or in groups. Mapping can be done on paper, on computers, or by arranging note cards. Other variations exist; see

Ruiz-Primo and Shavelson (1996) and Shavelson, et al. (1994) for a discussion of many of these variations when maps are used for assessment in science education.

Once students have drawn maps, these maps can be scored quantitatively based on their characteristics (e.g., number of correct propositions, levels of hierarchy, cross links). Many scoring schemes exist; for a few examples, see Liu (1994) and Lomask, Baron, and Grieg (1993), Novak and Gowin (1984), and Novak and Musonda (1991).

There are several advantages to the use of map generation in the classroom or in research as a measure of connected understanding. The research is clear that many students learn from constructing maps. Maps actually look like they measure connected understanding; in fact, among the various kinds of assessments of structural knowledge, maps constructed by students are considered to be the most direct measure.

As Ruiz-Primo and Shavelson (1996) and Shavelson et al. (1994) suggested, the existing evidence of the validity of map (specifically, concept map) generation assessments is far from complete or persuasive. However, as with the set of traditional structural knowledge assessment techniques, there is growing research evidence of the validity of concept mapping as a measure of structural knowledge. See, for example, Markham, Mintzes, and Jones (1994) and Wallace and Mintzes (1990) for validity evidence.

Map generation has limitations as a classroom outcome assessment. First, as Shavelson et al. (1994) stated, there is no universally accepted and simple scoring system for concept maps. Second, maps are idiosyncratic since there are many correct (and, of course, incorrect) ways to characterize any set of concepts and their interrelationships; in addition, each mental network has some unique components. Third, students must learn to draw concept maps, a process that is time-consuming and can be tedious and frustrating. Some students (and instructors) do not like to draw concept maps (see, e.g., Anderson & Huang, 1989; Barenholz & Tamir, 1992). Fourth, map generation imposes a high level of cognitive demand, both spatial-visual and verbal, on students. This high demand level may be most problematic with lower achieving students and students whose first language is not the language used in the assessment task. The most important uses of these maps may be for dynamic, rather than outcome, assessment purposes; in dynamic assessment there is no need for a quantitative score.

Fill-in Maps

A second, rarely explored, mapping approach to assess student's connected understanding is through the use of fill-in maps. The general assessment process involves constructing a master map. Keeping that map structure intact, some or all of the concept words (and/or linkages) are omitted. Students fill in these blanks either by generating the words to use (called "generate-and-fill-in") or by selecting them from a set which may or may not include distractors (called "select-and-fill-in"). The selection set may be listed on the map itself. Surber (1984) may have been the first to use fill-in concept maps as an assessment approach. Naveh-Benjamin, Lin, and McKeachie (1995) also used a fill-in approach with hierarchical maps containing unlabeled links.

Purpose

This article presents research in which we developed and began to explore the validity of the fill-in concept map format as an assessment of middle school students' understanding of the relationships among selected science concepts and processes. Our assessment format was designed to complement, not replace, traditional and alternative types of assessments of achievement used in school-based research. In addition, we wanted a format that did not exhibit at least some of the limitations associated with map generation and with the traditional research approaches. Ideally, our measure:

- could be administered efficiently (in less than one class period) to a whole class;
- would not require computers for administration or scoring;
- would show students a visual structure as they completed the assessment;
- would contain labeled nodes and links since both are critical in connected understanding; and
- would yield at least one valid interpretable score that explicitly quantifies connected understanding.

Phase I of our work included the development and field testing of a variety of possible concept mapping formats. Phase II involved the creation, field-testing, and revision of the most promising format, the select-and-fill-in concept map, into its final form. In Phase III, we tested this format with ethnically diverse middle school science students.

We emphasized common validity concerns throughout our development and testing work. Specifically, we began to address concerns related to the map assessment content, cognitive processes required to complete the map assessment tasks, technical quality, and relationships of map scores and patterns of map scores to those from other assessment measures.

Participants

The teachers who worked with us were or had been participants in a teacher enhancement program called TOPS (Teacher Opportunities to Promote Science) designed and offered by Los Alamos and Sandia National Laboratories. This three-year program was designed to reach cohorts of 20-25 rural New Mexican middle school science and mathematics teachers. Its objectives were broad and included increasing teachers' knowledge of science and mathematics and of the applications of science and mathematics at the Laboratories and enhancing their teaching skills (partly through providing them with hands-on science and mathematics materials). The Program also included a leadership component.

The TOPS science teachers and their seventh and eighth grade students participated in this project. In addition, these teachers recruited colleagues from their schools who also taught seventh or eighth grade science. These students self-identified as members of one of three ethnic groups: Hispanic American, Native American, and White American. Students were volunteers and could stop participating at any time.

Phase I: Development and Selection of the Map Format

Phase I included three aspects. First, several different map assessment formats were created. Second, these were evaluated through group work with students. Third, the two most promising formats were selected and evaluated with individual students in an informal "think-aloud" protocol design. During this Phase, we emphasized validity concerns related to the process aspect of assessment content and the cognitive processes used by students to complete the assessment tasks.

Format Development

Based on an initial group discussion with TOPS teachers, concept maps were selected as the basic form for the map assessment. To begin to identify possible useful formats, we developed 25 different map assessment formats. These formats varied systematically based on two primary characteristics: the type of response required from the student to complete the map and the design of the map itself. All of the formats required students to respond either by selecting answers (with or without distractors) or by generating their answers. In three of the selection formats, students were provided with "puzzle" pieces that they could move around on their maps. Refer to Table 1 for a summary of the formats we tried.

Insert Table 1 about here

Group Evaluation

During the regular science periods in two TOPS teachers' classrooms, each of us worked with students in groups of three to six; about 210 students participated. After introducing ourselves and the purpose for our project to the class as a whole, we introduced the idea of concept maps, using a simple example. Students then viewed a portion of a video on geothermal energy to ensure that they had been exposed to the information required for potential successful completion of their map assessment formats. Groups were formed, and students and leaders completed a simple example that matched their format. Next, each student completed the format independently and silently (although there often was collaboration among group members) and then discussed their responses as a group. Finally, the students and teachers were thanked for their help.

We used quantitative results and qualitative observations from the group work to identify a few mapping formats for further study. All formats that included distractors were eliminated; if students chose one or more distractors early in completing their maps, these choices often made the rest of the map impossible to complete. The map designs missing either all nodes or all links, as well as the format that required generation of the map structure, also were eliminated as excessively difficult given the time constraint of one class period. As would be expected, consecutively missing nodes, links, or node-link combinations were more difficult than non-consecutive ones and so were eliminated. Designs that included missing links also were eliminated since many linkage words and phrases are quite general and so often fit several places in a map.

Although the "generate" response demand has good potential for classroom use by teachers, it was eliminated because it has three disadvantages for use in large-scale research projects. First, it depends heavily on the students' communication skills, including vocabulary, spelling, and handwriting. Second, it is not possible to identify all possible correct responses before students complete the assessment, making scoring difficult. Third, different scorers will disagree about what responses should be counted as correct; a group of experts would have to make these decisions after a list of all responses to each blank in the map had been tallied.

We also reluctantly eliminated the movable pieces formats from further consideration. Large numbers of sets were time-consuming to create and attaching the movable pieces to the maps was difficult. Like some of the other formats we tested, we believe that this format is promising for use by classroom teachers.

Individual Evaluation

In the group phase of the research, students seemed to approach the task of doing these map assessments in many different ways, even when given identical instructions. To pursue this finding and to continue to explore the potential use of map formats, we worked with 12 students individually in a third TOPS teacher's school. Based on the group results, two map designs with non-consecutive missing nodes were selected for continued testing; task demands included selecting responses from a list or from movable pieces. Although we had eliminated the movable pieces design from serious consideration in our study, we included it as a comparison for the selection design. The procedure was similar to that employed in our group work except that each student was asked to "think out loud" while completing the format and then to verbally answer questions about thinking processes, difficulty level of the task, selected correct and incorrect answers, and vocabulary.

Our results yielded four important outcomes. First, different students did approach these assessment differently. Because of multiple linkages in students' cognitive networks, we believe that there should be multiple strategies that result in correct responses. Some of the students looked at their maps as a whole and concentrated on relationships as they worked. One of these students indicated that he "talked it out in my head to make sure it made sense." This general type of strategy seemed to be the best. Alternatively, some looked at their maps in pieces. At the extreme of this strategy, one student worked "backwards" by selecting a response from the list and then trying to locate its position on the map. This general strategy type yielded poor results. Some began with the "easy ones first." Some started with what they learned first. Others started at a particular point in the map and followed the arrows. Some combined more than one of these approaches.

Second, students accepted the assessment task. Relevant student comments included: "It was fun;" "It was a good way to learn;" "It's a better way to test." Third, as is true with any assessment, some students missed answers because they did not have a good understanding of the vocabulary.

Fourth, we could see no differences in how students approached the movable pieces format and the select-and-fill-in format. These observations, in conjunction with our group work with students, suggested that the best approach to chose was the select-and-fill-in map format with up to half of the non-consecutive concept words removed and listed in a selection set.

Phase II: Development, Field-testing, and Revision of the Assessment Packet

Phase II included three parts. First, we developed draft versions of the measures we wanted to use. Second, we field tested these. Third, we revised them and assembled them into a test packet. During this Phase, we emphasized validity concerns related to content domains and technical quality; we also created the external measures to be used for comparison to our map measure.

Draft Measures

As part of our validity research, we wanted to compare students' performance on a traditional measure of science achievement with their performance on our fill-in concept maps. We selected multiple-choice items to comprise the traditional measure since this format has been used for decades to measure science achievement in large-scale standardized assessments. Good multiple-choice items measure understanding of science concepts; selecting correct answers may require students to access limited portions of their cognitive networks. We initially designed drafts of both of our achievement measures such that each assessed understanding in the same four major areas often covered in middle school science classes (life sciences, earth sciences, physical sciences, and scientific inquiry).

The first draft of the multiple-choice measure consisted of 40 items selected from the items released to the public that were used in the 1990 National Assessment of Educational Progress (NAEP) for students in the eighth grade. Students chose the best answer from a set of four options given for each item. NAEP items were used because these items were carefully constructed with input from relevant groups and materials from across the U.S.. According to NAEP information, each of the items could be classified as assessing some aspect of understanding in the life sciences, earth sciences, physical sciences, or scientific inquiry.

We constructed four concept maps, one to cover aspects of each of these same four areas. The life sciences were represented by a map about plants, the physical sciences by a map about energy, the earth sciences by a map about the earth, and scientific inquiry by a map about the nature of scientific knowledge.

We created the draft select-and-fill-in concept maps by removing 38 nonconsecutive concepts from this set of four maps; these words or phrases were placed in a corner box of the appropriate map. Students completed this assessment format by selecting a response from the box and writing it in the blank. About half of these fill-in nodes assessed connected understanding of the same concepts assessed in half of the multiple-choice items.

We included nine Likert scale items designed to measure selected aspects of students' attitudes toward science including, for example, self-efficacy, affect, and value. The 5-point response scale ranged from "Strongly Disagree" through "Neither" to "Strongly Agree." Most of these items were selected from the Longitudinal Study of American Youth. We could not include more attitude questions due to the need to complete data collection from each class within one class period.

Pilot Field Tests

We engaged in two kinds of pilot field testing. First, we validated the adequacy and coverage of the four concept maps using both teachers and discipline experts. Five TOPS middle school teachers evaluated all four maps for accuracy and completeness. They also indicated which map propositions matched and did not match each multiple-choice item. Four University faculty also evaluated the maps that assessed disciplines in which they were expert; each map was evaluated by two different faculty members, with some faculty evaluating more than one map.

Second, and concurrently, we pilot tested the two measures on 63 seventh and eighth grade students in four science classes taught by two TOPS teachers. About half of the students were female and half were male; similarly, Native American, Hispanic American, and White American students were about equally represented. Although almost all of the students seemed to understand the tasks and appeared capable of responding to them, it was difficult for them to complete all measures in some of the classes which had shorter class periods.

Final Measures

The draft versions of the two achievement measures were revised a number of times based on the results of the two kinds of field testing. These revisions primarily involved redrawing parts of the concept maps and eliminating multiple-choice items. The TOPS teachers had matched 12 multiple-choice items that were supposed to be unmatched items to the "Inquiry" map. Because of this outcome and of the need to shorten administration time, several multiple-choice items were eliminated. The final version of the map measure included 37 select-and-fill-in nodes; half (19) of these assessed connected understanding of concepts found on the multiple-choice assessment. The final version of the multiple-choice measure consisted of 27 items; 70% (19) of these tested concepts were also assessed in the map measure. Using responses from our pilot sample of students, we analyzed the internal consistency of each of these pilot measures using Cronbach's alpha. Values of Cronbach's alpha were .94 for the map measure and .85 for the multiple-choice measure. All items in both measures functioned adequately in their respective total score; the alpha values could not be increased by more than .01

through elimination of any of the items. We could not evaluate the internal consistency of the draft attitude measure because many of the pilot test students were unable to complete it due to time constraints. The final version of our packet included the two kinds of achievement formats as well as the nine-item attitude measure. See Figure 2 for an example of one of the four maps.

Insert Figure 2 about here

Each set of achievement items was contained in a test packet. For the multiple-choice items, students read each item and marked their response on a machine readable answer sheet. Students completed the map measure by writing directly in the map test packet. We later coded their responses and added them to the student's answer sheet. The attitude items and responses also were contained on the answer sheet.

Phase III

During Phase III, we tested the map measure with groups of rural culturally diverse middle school science students. We emphasized validity aspects related to technical quality and relationships with other measures.

Subjects and Data Screening

We examined our original sample and their responses to our measures. We dealt with missing data and outlying scores and then characterized our final analysis sample.

Original Sample

Two-hundred sixty-four Hispanic, Native American, and White students of four seventh grade teachers and 413 students of six eighth grade science teachers from five rural New Mexican schools participated in the study. At each grade level, about half reported that they were female (seventh: 48%; eighth: 52%). In the seventh grade, 18% reported that they were Hispanic, 47% Native American, and 36% White; in eighth grade, 21% reported Hispanic, 36% Native American, and 43% White.

Non-participants

It is common for middle school students to leave items blank. Average scores on the attitude survey were used so that students who had left some items blank could be included in the analyses. As is common with scoring achievement measures, blanks on the achievement measures were scored as incorrect. However, as students who did not complete enough of a measure to warrant computing a score as defined above could have inappropriately impacted results, these students were called non-participants. Non-participation occurred for a variety of reasons: for example, the student was called from the room in the middle of data collection, "gave up" on the measure, ran out of time, or forgot the last page.

To define non-participants, we examined each measure and logically set rules based on the number of completed responses we thought were required to compute reasonably accurate scores. In general, non-participants were defined as students who did not complete at least 75% of the measure.

Achievement non-participants. We used the 75% completion rule for the multiple-choice measure. For the map measure, we used this same rule with an extension. The maps were of differential difficulty, with maps 3 (scientific inquiry) and 4 (physical science) most difficult. In order to compute a meaningful score with blanks scored as incorrect, we identified students as non-participants if they failed to complete at least 75% of all items and/or left an entire map blank. Non-participants on either achievement measure were eliminated from all achievement analyses.

In the seventh grade, 26 students were non-participants on one or both achievement measures. This group included equal numbers by gender. However, ethnic percentages were not in proportion to their representation in the sample. Hispanics constituted 42% (11) of the non-participants, Native Americans 38% (10), and Whites 19% (5). Hispanics were over-represented while Whites were under-represented. Eliminating these non-participants left 238 seventh grade students in the achievement sample.

In the eighth grade, 16 students were achievement non-participants. Again, boys and girls split equally. In this grade, ethnic percentages were in approximate proportion to their representation in the sample: Hispanics constituted 25%, Native Americans 31%, and Whites 44%. Eliminating these non-participants left 397 eighth grade students in the achievement sample.

Attitude non-participants. Three seventh grade students (two White males and one Native American female) and six eighth grade students (three Hispanic males and three Hispanic females) did not complete at least 75% of the attitude items. This process left 261 seventh grade and 407 eighth grade students for the attitude analyses.

Outlying Scores

Scores that fall far from cell means can have undue influence on results. We examined score distributions in the most complex interactions cells (gender by ethnicity) to identify outliers. Scores that fell over 3 standard deviations from their cell means and were discontinuous from their closest neighboring scores were deemed outliers. They were eliminated and their cells checked again until no more outliers were identified. Outliers were eliminated from all subsequent analyses, with the exception of the analyses of technical quality (see below).

Achievement outliers. In seventh grade, two students were achievement outliers. One White male scored low on the maps measure, and one Native American female scored high on the multiple-choice measure. These students were eliminated. This process left 236 seventh grade students in the achievement analysis sample.

In eighth grade, two students were achievement outliers. Both were White females. One scored too low on the maps measure and the other scored too low on both measures. Eliminating them left 395 eighth grade students in the achievement analysis sample.

Attitude outliers. One White male in seventh grade and one Hispanic male in eighth grade scored low on the attitude survey. They were eliminated, leaving 260 seventh grade and 406 eighth grade students in the attitude analysis sample.

Analysis Sample

Eliminating non-participants and outliers gave us a final analysis sample for the achievement measures of 236 seventh grade students and 395 eighth grade students. As desired, the demographic characteristics at both grade levels remained about the same as those found in the original sample. At each grade level, about half reported that they were female: seventh grade - 48% (113); eighth grade - 51% (203); half reported that they were male: seventh grade - 52% (123); eighth grade - 49% (192). In the seventh grade, 15% (36) reported that they were Hispanic, 48% (112) Native American, and 37% (88) White. In eighth grade, 21% (81) reported Hispanic, 37% (145) Native American, and 43% (169) White. Demographic characteristics for the attitude analysis sample were similar.

Procedures

Either one or two people collected data in each classroom. We balanced the order of administration of the achievement measures. We randomly split each class in half; about half of the students completed the multiple-choice measure first followed by the fill-in concept map measure while the other half received the measures in the reverse order. The multiple-choice measure was shorter than the fill-in concept map measure; its format also was more familiar to the students. Thus, most students completed it more quickly than the fill-in concept map measure. To balance the time requirements, students completed the attitude items after finishing the multiple-choice items. The response form also asked them to report grade level, class period, gender, and ethnicity (Anglo/White, Hispanic, American Indian, African American, Other).

Results

The results from Phase III are presented in two sections. The first section reports findings from the analyses for technical quality. The second presents findings regarding score and pattern relationships with other measures and with student characteristics.

Technical Quality

We needed to evaluate our measures for internal consistency and item functioning before we could form scores to search for statistical outliers. For these analyses only, outliers were included; as with all analyses, non-participants were eliminated.

Since we analyzed our data separately for seventh and eighth grade students, we examined technical quality as assessed by traditional item analysis techniques separately by grade level also. Responses to all items in both achievement measures were scored as correct or incorrect before analysis. Responses to negatively-worded attitude items were reversed.

For the seventh grade, reliability analysis of the map measure yielded a Cronbach's alpha value of .92; for eighth grade, Cronbach's alpha was .91. For seventh grade, reliability analysis of the multiple-choice measure yielded a Cronbach's alpha value of .81; for eighth grade, Cronbach's alpha was .79. All items in both measures at both grade levels functioned adequately in their respective total score. Students were given percent correct map and multiple-choice scores for the remainder of the analyses.

At each grade level, the same two attitude items correlated negatively with total score. Deleting these items yielded an acceptable alpha value of .73 for the seventh grade and .76 for the eighth grade. Students were given a mean attitude score for the remaining attitude analyses. Higher scores meant more positive attitudes.

Score and Pattern Relationships with External Measures

We examined three sets of results related to score and pattern relationships. First, we examined mean map differences by grade level. Second, we correlated scores on the map measure with scores on the multiple-choice achievement measure and on the science attitude measure, a non-achievement measure. Third, we examined patterns of group mean differences on the map and multiple-choice measures to identify similarities and differences.

Grade Differences

If the map measure assessed science achievement, we would expect the mean score for eighth grade students to be higher than that for seventh grade students. We found this result. The mean map score in the eighth grade (66% correct) was eight percent higher than the mean map score in the seventh grade (58%).

This pattern matched that found with the multiple-choice scores. The mean multiple-choice score in the eighth grade (60% correct) was six percent higher than that in the seventh grade (54%).

Score Relationships

We included the multiple-choice measure as a comparison measure for achievement. Like others who research alternative assessments (e.g., Koretz, Stecher, Klein, & McCaffrey, 1994) we hypothesized a moderate to moderately high relationship between scores on a traditional measure of science achievement utilizing a multiple choice format and on our alternative assessment utilizing the select-and-fill-in concept map format because both are measures of science achievement. A relationship that is too low would indicate that our measure is not assessing science achievement; a relationship that is too high would indicate that our measure is assessing the same kind of knowledge as that assessed by multiple choice items.

The Pearson Product Moment correlation coefficient related total maps score to total multiple-choice score was .74 in the seventh grade and .77 in the eighth grade. These two score distributions shared 55% of their variance in the seventh grade and 59% of their variance in the eighth grade. Clearly these two measures assessed some of the same aspects of science understanding. However, there also was some unique variance, suggesting that they may also have assessed some different aspects of science understanding.

We included the attitude survey as a non-achievement comparison measure. Maps and attitude scores correlated .26 in seventh grade and .36 in eighth grade. These values were very similar to those for multiple-choice and attitude scores (as would be expected); for seventh grade, the correlation was .25 while for eighth grade it was .37.

Score Patterns

We examined similarities and differences in mean scores by gender and by ethnicity for the two achievement measures separately by grade level. Because order of administration could have affected the scores, it also was included as a factor in our three-way factorial design. All interactions were included in the model. The unique approach was used to accommodate the small degree of non-orthogonality among our factors; as appropriate for this approach, unweighted means were used to interpret significant effects. Tables 2 and 3 contain the ANOVA source tables and the relevant unweighted means. The patterns comparing significant effects for maps and multiple-choice are briefly described below.

Insert Tables 2 and 3 about here

As the seventh grade ANOVA source tables show, ethnicity was a significant effect for both maps and multiple-choice scores, although it was mediated by gender for multiple-choice scores. Examining the unweighted means indicates that Whites, on average, scored highest and Native Americans lowest on both measures. The exception to this pattern occurred for Hispanic males who scored at about the same level as Native American males.

The eighth grade ANOVA source tables show a similar pattern; ethnicity was the only significant effect. The unweighted means again show that, on average, White scored highest and Native Americans lowest.

Discussion

Our findings provide initial evidence that the select-and-fill-in concept map assessment format can be a valid measure of science achievement when used with rural, ethnically diverse middle school students. We were able to begin to address four important aspects of validity: content, cognitive processes, technical quality, and score relationships.

Public school teachers and academic subject experts agreed that the maps were accurate and covered the topics. In addition, our work with individual students suggested that they used connection strategies to complete the maps. Item analyses indicated that the maps possessed high internal consistency, providing some evidence of technical quality.

Students take science throughout middle school. If our measure assessed science achievement, the eighth grade mean map score should have been higher than the seventh grade mean score; it was. We also hypothesized that the map and multiple-choice score distributions should correlate moderately highly because they both measure science achievement of the same content; they should not correlate too strongly because we developed the map measure to assess connected understanding, a type of understanding that is difficult to assess with a multiple-choice format. We found this pattern to some extent. At each grade level, the two score distributions correlated in the middle .70's, providing strong evidence for convergent validity (approximately half of their variance shared). Even so, there was some evidence of divergent validity since the other half of their variance was unique. The relationship between map and attitude scores was lower, as it should be, providing additional divergent validity evidence. In addition, we found similar similarities and differences in score patterns on both achievement measures across gender and ethnic groups.

The best achievement comparison measure would have been another assessment of connected understanding. However, the TOPS teachers were not interested in either the traditional research-based kinds of assessments (e.g., relatedness ratings) or in having their students draw concept maps. Thus, we were unable to administer either of these kinds of measures. However, we currently are working with undergraduate students enrolled in introductory astronomy courses and with their instructors. We have administered a select-and-fill-in concept map measure and a relatedness ratings measure assessing astronomy understanding. The resulting correlation was .50, providing both convergent and discriminant evidence in regard to assessing structural knowledge.

The development of an efficient, valid means of assessing connected understanding will be an invaluable tool for science educators and researchers. We believe that we have taken some first steps toward achieving this goal. Initial evidence indicates that the select-and-fill-in concept map format measures connected understanding in ethnically diverse middle school and undergraduate science students.

Table 1: Summary of map designs and response demands explored in Phase I

Response demands:	Select	Select (with (distractors)	Move	Generate
Missing map elements:				
All nodes	yes	yes	yes	yes
All links	yes	yes	—	yes
50% nodes	yes	yes	—	yes
50% links	yes	yes	—	yes
Non-consecutive nodes & links	yes	yes	—	yes
Some consecutive nodes & links	yes	yes	yes	yes
Entire Structure	yes	—	yes	—

Table 2. Achievement Results for Seventh Grade Students

Maps Measure:	SS	df	MS	F	p
Gender	185.73	1	185.73	.51	.47
Ethnicity	31170.32	2	15585.16	43.17	.00
Order	38.19	1	38.19	.11	.75
Gender by Ethnicity	475.72	2	237.86	.66	.52
Gender by Order	111.00	1	111.00	.31	.58
Ethnicity by Order	496.85	2	248.43	.69	.50
Gender by Ethnic by Order	504.90	2	252.45	.70	.50
Error	80869.45	224	361.02		
Total	115288.98	235	490.59		

Multiple-choice measure:	SS	df	MS	F	p
Gender	42.76	1	42.76	.17	.68
for Whites	667.58	1	667.58	2.71	.29
for Hispanics	947.29	1	947.29	3.85	.06
for Native Americans	143.39	1	143.39	.58	.82
Ethnicity	24022.16	2	12011.08	48.82	.00
for females	9925.95	2	4962.98	20.17	.00
for males	15982.12	2	7991.06	32.48	.00
Order	77.95	1	77.95	.32	.57
Gender by Ethnicity	1600.06	2	800.03	3.25	.04
Gender by Order	33.85	1	33.85	.14	.71
Ethnicity by Order	1163.91	2	581.95	2.37	.10
Gender by Ethnic by Order	3.60	2	1.80	.01	.99
Error	55111.92	224	246.04		
Total	83685.74	235			

Unweighted Means:	Whites	Hispanics	Native Americans
Maps	72%	55%	47%
Multiple-choice	66%	51%	44%
for females	64%	57%	43%
for males	69%	46%	45%

Table 3. Achievement Results for Eighth Grade Students

Maps Measure:

	SS	df	MS	F	p
Gender	669.97	1	669.97	2.00	.16
Ethnicity	41698.08	2	20849.04	62.12	.00
Order	341.79	1	341.79	1.02	.31
Gender by Ethnicity	523.61	2	261.80	.78	.46
Gender by Order	288.42	1	288.42	.86	.36
Ethnicity by Order	1214.42	2	607.21	1.81	.17
Gender by Ethnicity by Order	314.44	2	157.22	.47	.63
Error	128543.98	383	335.62		
Total	173165.89	394			

Multiple-choice Measure:

	SS	df	MS	F	p
Gender	681.99	1	681.99	2.90	.09
Ethnicity	25076.11	2	12538.05	53.34	.00
Order	223.09	1	223.09	.95	.33
Gender by Ethnicity	191.69	2	95.85	.41	.67
Gender by Order	347.12	1	347.12	1.48	.23
Ethnicity by Order	385.52	2	192.76	.82	.44
Gender by Ethnicity by Order	956.16	2	478.08	2.03	.13
Error	90023.36	383	235.05		
Total	118098.49	394			

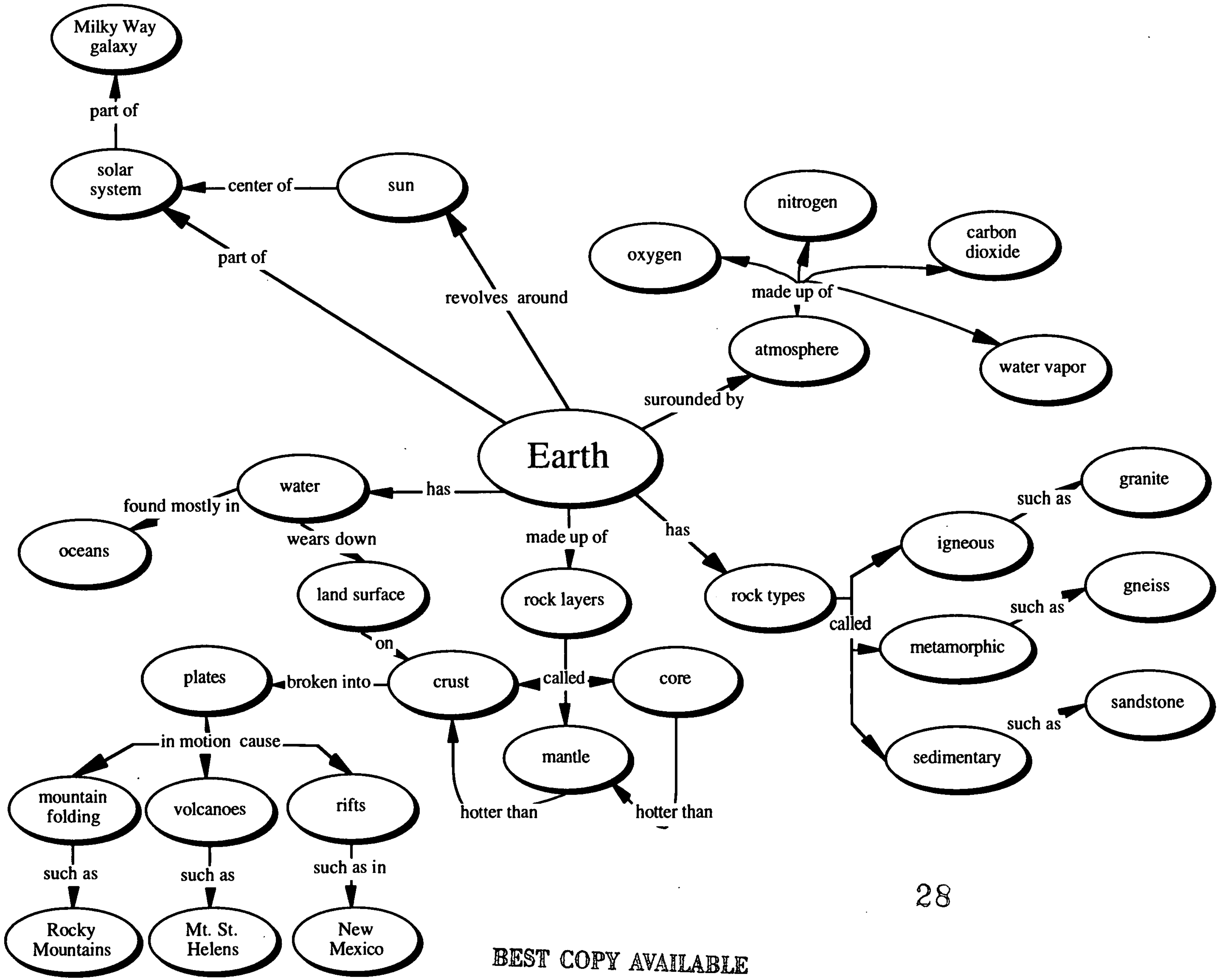
Unweighted Means:

	Whites	Hispanics	Native Americans
Maps	77%	67%	53%
Multiple-choice	70%	59%	52%

Figure Captions

Figure 1. Example concept map representing aspects of earth sciences

Figure 2. Select-and-fill-in concept map example based on Figure 1



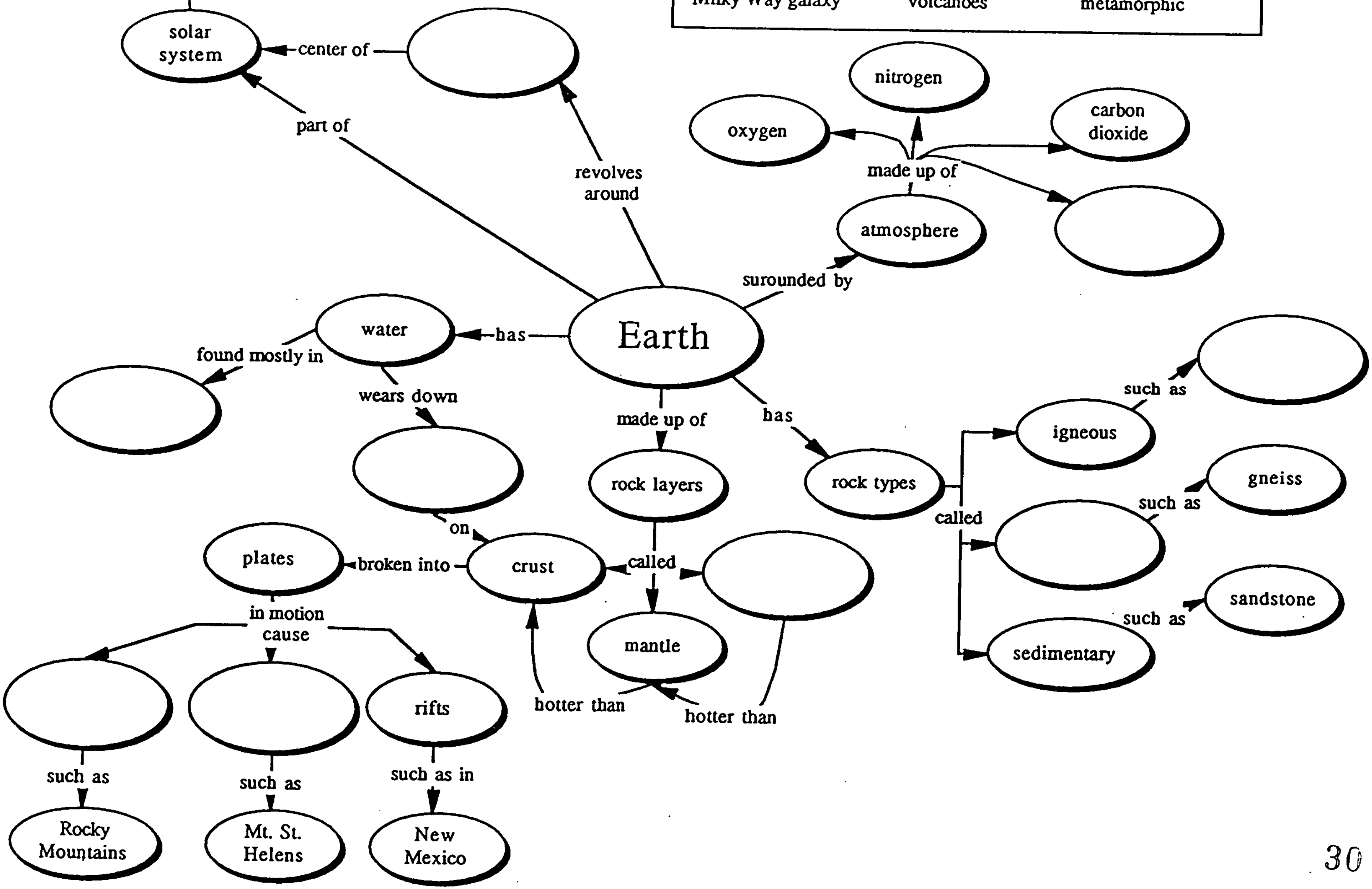
28

BEST COPY AVAILABLE

MAP 2

Answer list for Map 2

mountain folding	water vapor	granite	sun
land surface	oceans	core	
Milky Way galaxy	volcanoes	metamorphic	



BEST COPY AVAILABLE

References

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. Journal of Educational Psychology, 86, 303-311.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. New York: Oxford University Press.

Anderson, T. H., & Huang, S. C. C. (1989). On using concept maps to assess the comprehension effects of reading expository text. (Report No. 483). Champaign, IL: University of Illinois (ERIC Document Reproduction Service No. ED 310 368).

Barenholz, H., & Tamir, P. (1992). A comprehensive use of concept mapping in design instruction and assessment. Research in Science and Technological Education, 10, 39-52.

Goldsmith, T. E., & Johnson, P. J. (1990). A structural assessment of classroom learning. In R. W. Schvaneveldt (Ed.), Pathfinder associative networks: Studies in knowledge organization (pp. 241-254). Norwood, NJ: Ablex.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. Journal of Educational Psychology, 83, 88-96.

Johnson, P. J., Goldsmith, T. E., & Teague, K. W. (1995). Similarity, structure, and knowledge: A representational approach to assessment. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), Cognitively diagnostic assessment (pp. 221-249). Hillsdale, NJ: Erlbaum.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. Hillsdale, NJ: Erlbaum.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont Portfolio Assessment Program: Findings and implications. Educational Measurement: Issues and Practice, 13 (3), 5-16.

Liu, X. (1994, April). The validity and reliability of concept mapping as an alternative science assessment when item response theory is used for scoring. Paper presented at the meeting of the American Educational Research Association, New

Orleans, LA.

Lomask, M. S., Baron, J. B., & Grieg, J. (1993). Assessing conceptual understanding in science through the use of two- and three-dimensional concept maps. Paper presented at the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Ithaca, NY.

Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). The concept map as a research and evaluation tool: Further evidence of validity. Journal of Research in Science Teaching, 31, 91-104.

Marshall, S. (1995). Schemas in problem solving. Cambridge: Cambridge University Press.

National Research Council (1996). National Science Education Standards. Washington: National Academy Press.

Naveh-Benjamin, M., Lin, Y, & McKeachie, W. J. (1995). Inferring students' cognitive structures and their development using the "Fill-in-the-Structure" (FITS) technique. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), Cognitively diagnostic assessment (pp. 279-304). Hillsdale, NJ: Erlbaum.

Naveh-Benjamin, M., McKeachie, W. J., & Lin, Y. (1989). Use of the ordered-tree technique to assess students' initial knowledge and conceptual learning. Teaching of Psychology, 16, 182-187.

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y., & 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. (1984). Learning how to learn. NY: Cambridge University Press.

Novak, J. D., & Musonda, D. (1991). A twelve year longitudinal study of science concept learning. American Educational Research Journal, 28, 117-153.

Ruiz-Primo, M. A., and Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. Journal of Research in Science Teaching, 33, 569-600.

Shavelson, R. J., Lang, H., & Lewin, B. (1994). On concept maps as potential "authentic" assessments in science (CSE Tech. Rep. No. 388). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

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

Wallace, J. D., & Mintzes, J. J. (1990). The concept map as a research tool: Exploring conceptual change in biology. Journal of Research in Science Teaching, 27, 1033-1053.

AREA 1997

SE060390

ERIC

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Use of Fill-in Concept Maps to Assess Middle School Students'...	
Author(s): Candace Schau, Nancy Mattern, R. Weber, K. Minnick, C. Witt	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] ← Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document → [ ]

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here
Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Candace Schau	Position: Professor, Psychological Foundations
Printed Name: Candace Schau	Organization: College of Education
Address: Simpson Hall, University of New Mexico, Albuquerque, NM 87131	Telephone Number: (505) 277-3208
	Date: 4/1/97

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 21, 1997

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae2.educ.cua.edu.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (523)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:

AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.