THE CONCEPT MAP AS A RESEARCH TOOL: EXPLORING CONCEPTUAL CHANGE IN BIOLOGY

JOSEPHINE D. WALLACE
Department of Science Education, East Carolina University

JOEL J. MINTZES
Department of Biological Sciences, University of North Carolina at Wilmington

Abstract

This study examined the concurrent validity of concept maps as vehicles for documenting and exploring conceptual change in biology. Students (N = 91) who enrolled in an elementary science methods course were randomly assigned to one of two treatment groups. Subjects in both groups were administered a multiple-choice/free-response inventory which assayed their knowledge of "Life Zones in the Ocean," and then were asked to construct a concept map on the same topic. Those in the experimental group subsequently received 45 minutes of computer-assisted instruction on marine life zones, while those in the control ("placebo") group received an equivalent exposure to an unrelated topic ("Body Defenses"). Upon completing the instructional sequence, subjects were again administered the "Life Zones" inventory and asked to develop a postinstruction concept map on marine life zones. The data analysis employed a split plot factorial design with repeated measures. Differences among treatment groups were documented by analysis of variance and chi-square procedures. Subjects in the experimental group showed evidence of significant and substantial changes in the complexity and propositional structure of the knowledge base, as revealed in concept maps. No such changes were found in the control group. Results suggest that concept mapping offers a valid and potentially useful technique for documenting and exploring conceptual change in biology.

For the past decade we have focused our research efforts on issues of conceptual development and cognitive processes in biology (Mintzes, Trowbridge, Arnaudin & Wandersee, 1991). Several of these studies have employed cross-age designs in an attempt to understand how children's biological concepts change as they progress through the school years (Arnaudin & Mintzes, 1985; Trowbridge & Mintzes, 1985; 1988). Others have looked at the effects of brief interventions on student understandings in biology (Arnaudin, Mintzes, Dunn & Shafer, 1984). One ongoing study addresses differences in the content, structure, and use of biological knowledge among "novices" and "experts" (Maguire, 1990).
The common thread running through these studies is an interest in documenting and exploring conceptual change in the life sciences. In attempting to do so we have scoured the landscape of existing research tools and have used a wide variety of both qualitative and quantitative techniques, including clinical interviews (Pines, Novak, Posner & Vankirk, 1978), sorting tasks (Gobbo & Chi, 1986), open-ended and multiple-choice test items (Bloom, Hastings & Madaus, 1971), Likert-type (1932) scales, student drawings (Quiggin, 1977; Porter, 1974) and concept maps (Novak & Gowin, 1984). Of the approaches we have used, the combination of clinical interviews and concept maps has proven most versatile. But as researchers we have an obligation to be cautious and circumspect about any new investigative technique.

What of the concept map itself? How well do concept maps represent what students know and how they organize their knowledge? To what extent are changes in students' cognitive structures reflected in the concept maps they draw?

This study examined the concurrent validity (American Psychological Association, 1974; Anastasi, 1968) of concept maps as vehicles for documenting and exploring changes in cognitive structure. Specifically, it looks at the extent to which concept maps reveal changes in cognitive structure that result from brief episodes of instructional intervention. Should the concept map offer a valid mechanism for measuring conceptual change, we might reasonably suggest that it holds promise as a research tool and possibly as an adjunct or even an alternative to traditional evaluation methods.

**Concept Maps in Research**

A significant number of recent studies have examined the use of concept maps in science learning (Abayomi, 1988; Bar-Lavie, 1988; Bodolus, 1987; Cliburn, 1990; Curry, 1986; Heinze-Fry, 1987; Kurzmyn, 1987; Lehman, Carter & Kahle, 1985; Loncaric, 1986; Merrill, 1987; Novak, Gowin & Johansen, 1983). Most of these studies have focused on the use of concept mapping as a heuristic in promoting meaningful learning in a variety of knowledge domains and instructional settings. Others have employed the concept map as a metalearning strategy in an attempt to help students "learn how to learn." In general the results suggest that concept maps hold great promise, particularly as a way of "empowering" students through knowledge of their own learning.

Relatively few studies have investigated the use of concept maps as an approach to documenting and exploring conceptual change (Arnaudin et al., 1984; Bar-Lavie, 1988; Beyerbach, 1986; Stuart, 1985). This is especially surprising in light of Carey's (1986) suggestion that concept maps be used to investigate knowledge acquisition. Says she, "By comparing successive concept maps, produced as the student gains mastery of the domain, the researcher can see how knowledge is restructured in the course of acquisition."

Early studies by Novak et al. (1983) reported low correlations between concept mapping scores and such conventional measures of learning as SAT scores and course examination grades. However, to our knowledge, the issue of validity of concept mapping as an evaluation approach has not been fully explored in a carefully controlled experimental environment. The present study sought to do just that, through random assignment of subjects to treatment groups, use of computer-assisted instruction to minimize variance attributable to instructional delivery, and careful documentation of learning by conventional psychometric testing.
THE CONCEPT MAP AS A RESEARCH TOOL

Method

Subjects

Subjects of this study were 111 elementary education majors who enrolled in five sections of a science education methods course at East Carolina University, Greenville, North Carolina. The course is normally taken during the junior year prior to student teaching and requires, as prerequisites, a physical science and a life science course for elementary school teachers.

Of the original 111 subjects, 20 were lost to absenteeism and normal class attrition. As a result, complete data were available for 91 subjects (90 females and one male).

Procedure and Instrumentation

The entire experiment took place during six, 75-minute class sessions, and spanned a period of three weeks. One session focused on each of the following activities: training, practice, review, pretesting, instruction, and posttesting. Upon completion of the experiment, a series of follow-up interviews were conducted.

Training. The first session was devoted to training subjects in the concept mapping technique. In this period, subjects were introduced to operational definitions of terms applied to concept maps: concepts, propositions, principles, theories, relationships, hierarchy, cross-links, and general-to-specific examples. Subjects were then presented with examples of student-constructed concept maps, introduced to scoring criteria, and given examples of scored maps.

Practice. During the second session, subjects were given an opportunity to practice the concept mapping technique. Activities included a card-sorting exercise for generating concept maps (Novak & Gowin, 1984), a review and discussion of student-generated maps, and a homework assignment requiring subjects to develop and submit a concept map based on their reading of a selected chapter in a middle school science textbook.

Review. The third session was given over to review. The scored concept maps were returned to the subjects, the scoring method was reexamined, and problematic areas were discussed. Acetate transparencies of student-constructed maps served as models for discussion.

Pretesting. The focus of the fourth session was pretesting. During this period, subjects were administered a multiple-choice/free-response inventory which assayed their knowledge of the experimental software package, "Life Zones in the Ocean." The purpose of this pretesting and subsequent posttesting procedure was to document the extent of the learning that had occurred, using conventional psychometric methods. The testing instrument was developed by the software publisher, Prentice-Hall/Edunetics (1986a, 1986b), and consists of a set of 40 knowledge, comprehension, and application items. The alpha reliability was calculated at 0.76.

Following administration of the test, subjects were given a set of ten concept labels and asked to develop a preinstruction concept map on the topic, "marine life zones." The following concept labels were supplied to assist the subjects with the task: life zones in the ocean, abyssal zone, adapt, continental shelf, intertidal zone, mineral, neritic zone, photosynthetic zone, phytoplankton, and tide.

Instruction. The fifth session was devoted entirely to computer-assisted instruction (CAI). Using a table of random numbers, subjects were assigned to experimental and
control ("placebo") treatment groups. Subjects \( n = 42 \) in the experimental group engaged in a CAI program which introduced basic concepts of "Life Zones in the Ocean." Those in the control group \( n = 49 \) received an equivalent exposure to an unrelated topic: "Body Defenses." The programs, which are commercially available and widely used (Prentice-Hall/Edunetics, 1986a, 1986b), were chosen because they introduce domains of knowledge with which the subjects were judged to be relatively unfamiliar. The mean completion time for both programs is approximately 45 minutes. However, the subjects were encouraged to take as much time as they needed, to review portions of the program where necessary, and to take notes if they so desired.

The experimental program, "Life Zones in the Ocean," introduces students to environmental conditions and life in the intertidal zone, variation of abiotic factors (pressure, light, and temperature) within the neritic and open sea zones, and adaptations of plants and animals within these zones. The control program, "Body Defenses," examines the body's first, second, and third lines of defense against disease. Important concepts include mechanical and chemical barriers to infection, the role of white blood cells in fighting pathogens, and antigen/antibody reactions.

Posttesting. The sixth and final session provided a time for posttesting. During this period subjects were retested on their knowledge of "Life Zones in the Ocean" and were asked to develop a postinstruction concept map on the topic "marine life zones." The posttesting instrument was identical to that used as a pretest. To the extent possible, subjects in the experimental and control groups were treated identically in all respects.

Post Hoc Interviews

One week after the completion of the experiment, ten subjects in the experimental group were interviewed on their understanding of the program, "Life Zones in the Ocean." Five of these subjects had the lowest composite scores on the postinstruction concept maps; the other five subjects had the highest composite scores.

Twenty-two line drawings taken from the software program served as visual stimuli which helped to elicit propositional statements. The subjects were presented with each stimulus and prompted to, "Tell me everything you can think of about this picture from the unit 'Life Zones in the Ocean'."

The interviews were conducted in the privacy of a small room and responses were audiotaped. Subsequently the tape recordings were transcribed verbatim and divided into propositional statements.

Data Analysis

The data consisted of pre- and posttest scores on the "Life Zones" inventory, and scores on the preinstruction and postinstruction concept maps. The analysis employed a split plot factorial design with repeated measures (Campbell & Stanley, 1963; Pedhazur, 1982). Such a design enables the researcher to partial out variance attributable to differences among subjects, thereby reducing the error variance.

A preliminary analysis of scores on the "Life Zones" inventory revealed, as expected, small but significant differences \( F(1, 89) = 5.47, p < .05 \) favoring the experimental group (Table I). These differences, however, appear to be largely attributable to pretest differentials, and the measurable effect of the experimental treatment on the
posttest scores seems incremental at best. The analysis then shifted to the concept maps. To what extent are these achievement differences reflected in the concept maps that students draw? Two aspects of the concept maps were of interest: the concept mapping scores and the number of critical concepts and propositions depicted.

**Concept Mapping Scores.** Concept mapping scores reveal the complexity of the knowledge base and the structure of scientifically acceptable propositions embedded in it. The scoring categories include the number of relationships, levels of hierarchy, branchings, cross-links, and general-to-specific examples depicted in the concept map. Several different scoring and weighting schemes have been used to quantify these elements (Arnaudin et al., 1984; Novak, 1981; Novak & Gowin, 1984; Stuart, 1985). For the purposes of this study, the following scoring method was used:

- **Relationships:** one point for each valid proposition
- **Hierarchy:** five points for each valid level of hierarchy
- **Branching:** one point for each branching
- **Cross-Links:** ten points for each valid cross link
- **General to Specific Examples:** one point for each valid example.

**Critical Concepts and Propositions.** In an attempt to gauge the extent of biologically meaningful knowledge revealed in the concept maps, a panel composed of three outside biology educators was asked to review the program “Life Zones in the Ocean” and to develop their own concept maps. An analysis of these “template” maps produced a set of “critical concepts and propositions,” that is, a set of concepts and propositions which was deemed to be central to an understanding of the program.

Subsequently, each of the preinstruction and postinstruction maps was scored for critical concepts and propositions. One point was given for each critical concept and for each critical proposition or equivalent proposition judged to have the same semantic meaning. The critical concepts are listed below. (Critical propositions are identified in Figures 7, 8, and 9.)
Results and Discussion

Student Concept Maps

Figures 1 and 2 are examples of postinstruction concept maps drawn by students in the control and experimental groups. These particular concept maps, like all others, are idiosyncratic representations of domain-specific knowledge. Consequently, they are neither "representative" nor "typical" of any group. Nevertheless, as examples, the maps offer insight into some of the characteristics that are frequently seen in students with differing levels of conceptual understanding.

On careful observation it is readily apparent that Barbara's map depicts more valid relationships than does Jennie's. Furthermore, the hierarchical structure of Barbara's map is considerably more intricate, as is the branching pattern and the extent of cross-linkages, thus suggesting a substantially more differentiated and cohesive knowledge base. Even a cursory, qualitative examination of Jennie's map reveals a linear assemblage
of linked concepts that are poorly integrated with related concepts, traits often seen in concept maps of relative novices.

In addition to these traits, Jennie's map lacks any valid general-to-specific examples. Interestingly, both she and Barbara depict "high" and "low" as examples of the concept tide; however, both maps subordinate the latter concept to neritic zone, resulting in an invalid propositional hierarchy.

The number of critical concepts and propositions in Barbara's concept map is somewhat greater than that in Jennie's; however, an important difference is seen in the number of alternative propositions or "misconceptions" in the two maps. As might be expected, virtually all of Jennie's propositions are scientifically unacceptable. And, while Barbara does display an array of conceptual problems, a number of her propositions are consistent with the explanations given in the software program.

A qualitative analysis of student concept maps such as these can offer rich and detailed insights into the extent of meaningful learning resulting from classroom instruction. Such an analysis can also provide invaluable feedback for improving instruction.

Fig. 2. Barbara's concept map: an example from the experimental group.
For purposes of this study, however, the primary concern is the quantitative analysis of differences in the complexity and propositional structure of the knowledge base as depicted in student maps.

**Concept Mapping Scores**

Figure 3 summarizes the analysis of scores in the concept maps of students before and after instruction. The analysis of variance revealed significant differences among treatment groups in all five scoring categories.

The number of relationships in the concept maps of experimental subjects more than doubled as a result of instruction, while no substantive changes were found in the control group. These findings imply a significant shift in the degree of concept differentiation and a change in the total number of scientifically acceptable propositions encoded and stored in cognitive structure.

Additional evidence of progressive change in the knowledge structure is revealed in the levels of hierarchy and the extent of branching depicted. The hierarchy scores of the experimental group show an improvement of some 30%, in contrast to a slight decline in the control group. Somewhat similar results are seen in the patterns of branching where students in the control group actually improved by about 50% (possibly a "practice effect"), while those in the experimental group jumped by over 120%.

The number of cross-links shown in the concept map is taken as a measure of integration or cohesion in the knowledge base. Here we find a substantial amount of variability from map to map, with some students depicting several such links and others drawing none at all. Nevertheless, a comparison of pre- and postinstruction scores reveals no change in the control group and a severalfold improvement in the experimental group.

Finally, the number of general-to-specific examples depicted in the concept maps was relatively small and the extent of variability was fairly high, with a slight decline seen in both groups. Although significant differences were found favoring the experimental group, it is uncertain in this case whether these differences can be meaningfully interpreted.

**Critical Concepts and Propositions**

The mean number of critical concepts and propositions appearing in the concept maps is summarized in Figure 4. The summary data suggest no change in the control group but a substantial boost in the postinstruction maps of subjects in the experimental group, where the number of critical concepts increased by over two-thirds and the number of propositions jumped by nearly 150%. The analyses of variance confirm these interpretations, indicating significant differences favoring the experimental group in both critical concepts \([F (1, 89) = 36.33, p < 0.01]\) and propositions \([F (1, 89) = 50.24, p < 0.01]\).

A frequency analysis of student-generated critical concepts is given in Figures 5 and 6. These findings suggest substantial changes in the concept maps of subjects in the experimental group with correspondingly little change in the control group. In several instances \((C1, C2, C7)\), the frequencies of critical concepts rose from 0 to nearly 50%, in others \((C4, C5, C9, C10)\) the improvement was less dramatic. In only two instances \((C3, C11)\) were no differences detected.
Fig. 3. Effects of brief episodes of instructional intervention on concept mapping scores. $F(1, 89) = 30.3^{**}$ (relationships); $23.7^{**}$ (hierarchy); $29.8^{**}$ (branchings); $12.3^{**}$ (cross-links); $4.3^{*}$ (general-to-specific examples).
Fig. 4. Number of critical concepts and propositions in preinstruction and postinstruction concept maps.
Fig. 5. Frequencies of critical concepts (C1 to C7) generated by students in pre- and postinstruction concept maps. \( \chi^2 (3, N = 91) = 57.3** (C1); 49.8** (C2); 3.4 (C3); 24.3** (C4); 27.9** (C5); 19.5** (C6); 45.4** (C7) \).
CRITICAL CONCEPTS
GENERATED BY STUDENTS

EXPERIMENTAL (N=42)
CONTROL (N=49)

*P< .05
**P< .01

C8: PRESSURE
C9: FOOD CHAIN
C10: PRODUCERS
C11: CONSUMERS
C12: FISH
C13: ORGANISMS
C14: PHOTOSYNTHESIS

Fig. 6. Frequencies of critical concepts (C8 to C14) generated by students in pre- and postinstruction concept maps.

$[\chi^2 (3, N = 91) = 41.8** (C8); 16.4** (C9); 10.0* (C10); 5.9 (C11); 37.6** (C12); 11.6** (C13); 14.3** (C14)]$
In addition to these 14 critical concepts generated by the students themselves, 6 were supplied to the subjects to help them construct their maps. Except for one (adaptations), these concepts appeared in 83%–100% of the concept maps, with no differences between treatment groups. Adaptations appeared in the postinstruction maps of 49% and 90% of control and experimental subjects, respectively \( \chi^2 (3, N = 91) = 25.35, p < 0.01 \).

A similar but more complex pattern is seen in the frequencies of critical propositions (Figures 7, 8, and 9). Here significant and, in several cases, substantial differences were found in the frequencies of 17 out of 20 propositions. Except for one instance (P4) where prior differences were found, virtually all of the changes in the concept maps appear to be directly attributable to the brief instructional intervention.

For students in the experimental group, pre- to postmap improvement ranged from 17% (P15) to 50% (P6, P10). In the control group, differences ranged from a 14% decline (P17) to a 17% improvement (P1), with the frequencies of most propositions showing essentially no change at all.

In some instances the frequency of a proposition can be explained, in part, by the source and hierarchical locus of its constituent concepts. For example, the remarkably high frequencies of the first two propositions (P1 and P2) are instances where both concept labels were supplied and where the concepts appear high in the knowledge structure. On the other hand, as one progresses from “supplied” to “student generated” concepts, and from higher to lower levels in the hierarchy (P11 through P16), the frequencies of propositions decline substantially.

**Post Hoc Interviews**

Figure 10 summarizes the mean number of critical concepts and propositions in the interview transcripts of two groups of experimental subjects, those whose composite postinstruction mapping scores were the five lowest and the five highest. The results indicate that the highest exceed the lowest in critical concepts by about one-third (13.2 versus 10.0) and in critical propositions by almost 100% (58.8 versus 29.8).

These differences suggest a potentially strong relationship between the complexity of cognitive structure as revealed in concept maps and the extent of biologically meaningful knowledge possessed by the learner. Although we find these results strongly suggestive, the small sample size precludes further analysis. Additional studies which combine the use of concept maps with clinical interview methods in this way would be very helpful.

**Conclusions and Implications**

This study examined the concurrent validity of concept maps as vehicles for documenting and exploring conceptual change in biology. The results suggest that substantial and potentially important changes in both the complexity and propositional structure of the knowledge base are revealed in concept maps after only 45 minutes of computer-assisted instruction. Consequently, it appears that concept mapping offers a valid and useful mechanism for looking at changes in cognitive structure.

The findings of this study, along with some 10 years of practical experience in using concept maps, have convinced us that many researchers might profit by adding this approach to their personal repertoire of research methods. In using concept maps,
CRITICAL PROPOSITIONS 1 - 7

P1 LIFE ZONES IN OCEAN (I) INCLUDE INTERTIDAL (I)
P2 LIFE ZONES IN OCEAN (I) INCLUDE NERITIC (I)
P3 LIFE ZONES IN OCEAN (I) INCLUDE OPEN SEA (I)
P4 INTERTIDAL (I) LIES BETWEEN HIGH & LOW TIDE LINES (I)
P5 ORGANISMS (I) OF INTERTIDAL (I) ... EXPOSED TO ENV. CONDITIONS (I)
P6 ...TO LIVE IN INTERTIDAL (I) ORGANISMS ADAPT (I)
P7 EXAMPLES OF ADAPTATIONS (I) ARE... (I)

Fig. 7. Frequencies of critical propositions (P1 to P7) in pre- and postinstruction concept maps. \( \chi^2 (3, N = 91) = 7.8 \) (P1); 5.9 (P2); 27.8** (P3); 9.2* (P4); 33.7** (P5); 46.1** (P6); 27.1** (P7)
CRITICAL PROPOSITIONS 8 - 14

P8 NERITIC (⌀) LIES BETWEEN LOW-TIDE LINE (⌀) AND CONTINENTAL SHELF (⌀) EDGE
P9 OPEN SEA (舄) LIES BEYOND NERITIC (⌀) & INCLUDES OPEN OCEAN
P10 OPEN SEA (舄) INCLUDES PHOTOSYNTHETIC (舄), BATHYAL (舄) AND ABYSSAL (舄)
P11 ATMOSPHERIC PRESSURE (舄) IN OPEN SEA (舄) INCREASES WITH DEPTH
P12 LIGHT (舄) IN OPEN SEA (舄)...DECLINES WITH DEPTH
P13 TEMPERATURE (舄) IN OPEN SEA (舄)...DECLINES WITH DEPTH
P14 FISH (舄) ARE MOST ABUNDANT IN NERITIC (舄)

Fig. 8. Frequencies of critical propositions (P8 to P14) in pre- and postinstruction concept maps. $\chi^2 (3, N = 91) = 41.0** (P8); 53.5** (P9); 61.4** (P10); 37.4** (P11); 37.4** (P12); 27.1** (P13); 26.8** (P14)$
CRITICAL PROPOSITIONS 15 - 20

P15 ABUNDANCE OF FISH (0) DETERMINED BY AVAILABILITY OF FOOD PRODUCERS (0)
P16 IN FOOD CHAIN (0), PRODUCERS (0) EATEN BY CONSUMERS (0)... EATEN BY LARGER CONSUMERS (0)
P17 NERITIC (©) & PHOTOSYNTHETIC (©) ZONES SHARE FOOD CHAIN (0) OF
PHYOPLANKTON (©), ZOOPLANKTON (©) & FISH (0)
P18 NERITIC (©) HAS SECOND FOOD CHAIN (0) OF SEAWEED (0), SHRIMP (©) & FISH (0)
P19 FISH (0) ABUNDANT IN NERITIC (©) BECAUSE OF AVAIL. OF LIGHT (©) & MINERALS (©)
P20 VARIATIONS (©) IN FISH ARE ADAPTATIONS (©) TO ENV. CONDITIONS (©)

Fig. 9. Frequencies of critical propositions (P15 to P20) in pre- and postinstruction concept maps. $\chi^2 (3, N = 91) = 19.9^{**}$ (P15); $31.6^{**}$ (P16); $7.2$ (P17); $33.9^{**}$ (P18); $12.4^{**}$ (P19); $34.3^{**}$ (P20)
Number of critical concepts and propositions in interview transcripts.
we have found that they complement, but usually do not duplicate, the work of other techniques such as clinical interviews, sorting tasks, and conventional testing instruments. The concept map is the only approach we have found that attends to both what students know and how they organize their knowledge.

Of its many potential uses in research, we have found the concept map to be especially valuable in documenting the “intellectual journey” taken by students as they restructure their understandings in individual courses and across the school years. In our own work, for example, which explores the frequencies of “alternative conceptions” in biology (Arnaudin & Mintzes, 1985), we have enlisted the concept map in the exploratory, inductive, or qualitative phase of the work. In subsequent phases we have employed conventional psychometric approaches.

We should also add, from a purely practical standpoint, that concept mapping is quickly taught to experimental subjects, that it can be “administered” to large groups, and that the product is readily interpreted. These are no small considerations when one is attempting to probe understanding of complex, scientific concepts in young minds.

In closing, we wish to suggest, as many natural scientists have, that good research is theory driven but theory without method goes nowhere. Accordingly, we believe that good research in science education must rest on a firm foundation of learning theory; however, no matter how firm our theories are, understanding of complex educational events depends on the development and application of powerful new tools. We have found the concept map to be such a tool.

References


We wish to thank Ms. Lisa Pike for her graphics expertise.

Manuscript accepted August 21, 1990.