Concept Map-Based Assessment in Science:
Two Exploratory Studies

CSE Technical Report 436

Maria Araceli Ruiz-Primo, Susan Elise Schultz,
and Richard J. Shavelson
CRESST/Stanford University

August 1997

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-6511
(310) 206-1532

The work reported herein was supported in part under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA Catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education, and in part by a grant from the National Science Foundation (ESI-9596080, A002).

CONCEPT MAP-BASED ASSESSMENT IN SCIENCE: TWO EXPLORATORY STUDIES[1]

Maria Araceli Ruiz-Primo, Susan Elise Schultz,
and Richard J. Shavelson
CRESST/Stanford University

The search for new "authentic" science assessments of what students know and can do is well underway. As part of this search, educators and researchers are looking for more or less direct measures of students' knowledge structures. Concept maps—structural representations of key concepts in a subject domain, constructed by individuals—have been dubbed a potential "find."

The rationale behind this claim is that knowledge has an organizational property that can be captured with structural representations (e.g., Goldsmith, Johnson, & Acton, 1991; Jonassen, Beissner, & Yacci, 1993; White & Gunstone, 1992). Cognitive psychologists posit that "the essence of knowledge is structure" (Anderson, 1984, p. 5). Concept interrelatedness, then, is an essential property of knowledge. Indeed, one aspect of competence in a domain is that expert knowledge is well structured, and as expertise in a domain grows, through learning, training, and/or experience, the elements of knowledge become increasingly interconnected (e.g., Glaser & Bassok, 1989; Shavelson, 1972).

Assuming that knowledge within a content domain is organized around central concepts, to be knowledgeable in the domain implies a highly integrated conceptual structure. Concept maps, then, may capture important aspects of this interrelatedness between concepts.

A concept map is a structural representation consisting of nodes and labeled lines. The nodes correspond to important terms (standing for concepts)

---

[1] The authors are deeply grateful to Dr. Pinchas Tamir for his valuable comments on the two studies.

in a domain.[2] The lines denote a relation between a pair of concepts (nodes) and the label on the line tells how the two concepts are related. The combination of two nodes and a labeled line is called a proposition. A proposition is the basic unit of meaning in a concept map and the smallest unit that can be used to judge the validity of the relationship drawn between two concepts (e.g., Dochy, 1996). Concept maps, then, purport to represent some important aspect of a student's declarative knowledge in a content domain (e.g., chemistry).

Although the potential use of concept maps for assessing students' knowledge structures has been recognized (e.g., Jonassen et al., 1993; White & Gunstone, 1992), maps are far more frequently used as instructional tools (e.g., Briscoe & LaMaster, 1991; Holley & Danserau, 1984; Pankratius, 1990; Schmid & Telaro, 1990; Stice & Alvarez, 1987; Willerman & Mac Harg, 1991) than as assessment tools (but see, for example, Baxter, Glaser, & Raghavan, 1994; Beyerbach, 1988; Hoz, Tomer, & Tamir, 1990; Lomask, Baron, Greig, & Harrison, 1992).

Concept maps, as assessment tools, can be thought of as a set of procedures used to measure the structure of a student's declarative knowledge. We use the term "assessment" to reflect our belief that reaching a judgment about an individual's knowledge and skills requires an integration of several pieces of information; we consider concept maps as potentially one of those pieces (see Cronbach, 1990).

Ideally, before concept maps are used in classrooms or for large-scale assessment, and before concept map scores are reported to teachers, students, the public, and policy makers, research needs to provide information about the psychometric properties of concept maps for representing knowledge structure. Accordingly, the studies reported here provide evidence bearing on the reliability and validity of concept maps as representations of students' knowledge structures. In Study 1, we examine whether map scores are sensitive to who chooses the concepts to be used in the map (student or tester) and to the sampling of the concepts (e.g., random samples of key concepts from a domain). In Study 2, we examine whether traditional instructions to

---

[2] Actually, terms or words used in concept mapping are not concepts. They stand for concepts. Nevertheless, the terms used in concept mapping are called "concepts" and, from here on, we will follow this convention.

construct a hierarchical map are necessary, considering that the map should reflect the structure of the subject domain as represented in a student's memory rather than a preconceived psychological theory.

Concept Map-Based Assessment

Intuitively, the use of concept maps to evaluate students' declarative knowledge structure is appealing. A student's map construction directly reflects, to some degree, her or his understanding in a domain. Nevertheless, before adopting maps for assessment use, more needs to be known about them. A common understanding is needed of what a concept map assessment is and whether it provides a reliable and valid measure of a student's cognitive structure (Ruiz-Primo & Shavelson, in press).

A Concept Map-Based Measurement Framework

Concept map measures can be characterized by: (a) a *task* that invites a student to provide evidence bearing on his or her knowledge structure in a domain; (b) a *format* for the student's *response;* and (c) a *scoring system* by which the student's concept map can be substantively evaluated accurately and consistently. Without these three components, a concept map cannot be considered as a measurement tool (Ruiz-Primo & Shavelson, in press).

Based on this conceptualization, we found tremendous variation in what counted as concept mapping techniques. This variation in concept mapping techniques emerged from variations in task, response formats, and scoring systems (Ruiz-Primo & Shavelson, in press), and it is captured in Table 1.

Concept mapping tasks varied in three ways: (a) demands made on the students in generating concept maps (*tasks demands*), (b) constraints placed on the task (*task constraints*), and (c) the intersection of task demands and constraints with the structure of the subject domain to be mapped (*content structure*). As an example of the third category consider the constraint of building a hierarchical map in a subject domain that is not hierarchical. Methodologically and conceptually, there is no need to impose a hierarchical structure. If the content structure is hierarchical, and the student has mastered the domain, a hierarchical map should be observed.

Response formats vary in three ways: (a) whether the student's response is given with paper-and-pencil, orally, or on a computer (*response mode*); (b) the link between task and format (e.g., if the task asks the student to fill in a skeleton map, the response format provides the skeleton map; *response format*); and

Table 1

Concept Map Components and Variations Identified

Map assessment components	Variations	Instances
Task	Task demands	Students can be asked to: • fill-in a map • construct a map from scratch • organize cards • rate relatedness of concept pairs • write an essay • respond to an interview
	Task constraints	Students may or may not be: • asked to construct a hierarchical map • provided with the concepts used in the task • provided with the concept links used in the task • allowed to use more than one link between nodes • allowed to physically move the concepts around until a satisfactory structure is arrived at • asked to define the terms used in the map • required to justify their responses • required to construct the map collectively
	Content structure	The intersection of the task demands and constraints with the structure of the subject domain to be mapped.
Response	Response mode	Whether the student response is: • paper-and-pencil • oral • on a computer
	Format characteristics	Format should fit the specifics of the task
	Mapper	Whether the map is drawn by a: • student • teacher or researcher
Scoring system	Score components of the map	Focus is on three components or variations of them: • propositions • hierarchy levels • examples
	Use of a criterion map	Compare a student's map with an expert's map. Criterion maps can be obtained from: • one or more experts in the field • one or more teachers • one or more top students

Combination of map components and a criterion map	The two previous strategies are combined to score the students' maps.

(c) who draws the map (i.e., most frequently the student; however, teachers or researchers can draw maps from students interviews or essays; *mapper*).

Three general scoring strategies have been used with maps: (a) score the *components* of the students' maps (e.g., number of links); (b) compare the students' maps with a *criterion map* (e.g., a map constructed by an expert); and (c) a *combination of both strategies*.

If each of the six task demands (e.g., fill-in-the-blank nodes on a map) is combined with each of the eight types of task constraints (e.g., hierarchical vs. nonhierarchical), there are no less than 1530 (i.e., 6 x 2^8 - 1) different ways to produce a concept mapping task! Of course, not all combinations may be realistic. Regardless, the wide variety of potential maps raises issues about what is being measured. Some examples may help to make clear the problem of the variation in concept mapping techniques. Table 2 presents five examples of different types of tasks, response formats, and scoring systems used in practice and in research on concept maps (see Ruiz-Primo & Shavelson, in press, for more examples).

We suspect that different mapping techniques may tap different aspects of cognitive structure and may lead students to produce different concept maps. Nevertheless, current practice holds that all variations in mapping techniques are interpreted the same way, as representing aspects of a student's cognitive structure—the relationship of concepts in a student's memory (Shavelson, 1972). If concept maps are to be used as a measurement tool, we must take the time and effort to provide evidence on the impact of different mapping techniques for representing a student's knowledge structure.

Unfortunately, cognitive theory does not provide an adequate basis for deciding which technique to prefer because many of the techniques have no direct connection with one or another theory. Furthermore, current cognitive theories may be limited in their ability to guide mapping techniques because they tend to be middle-range theories focused on particular aspects of cognition. Application of cognitive theory, along with empirical research, should, over time, provide guidelines that would narrow the number of possible techniques to a manageable set.

In the meantime, research on concept maps should proceed by developing criteria from cognitive theory and practice that can help discard some techniques. In the studies reported here, we applied the following criteria to narrow down alternatives: (a) appropriateness of the cognitive demands required by the task; (b) appropriateness of a structural representation in a content domain; (c) appropriateness of the scoring system used to evaluate the accuracy of the representation; and (d) practicality of the technique. We eliminated, for example, a fill-in-the-blank task because we regarded it as inappropriate for measuring students' knowledge structures since the task itself too severely restricted the students' representations. We also favored scoring criteria that focused on the

Table 2

Five Examples of Different Types of Tasks, Response Format and Scoring Systems Used in Research on Concept Maps

Authors	Task	Response	Scoring system
Barenholz & Tamir, 1992	Select 20 to 30 concepts considered key concepts for a course in micro-biology and use them to construct a map.	Paper-and-pencil response. Students drew the concept map in their notebooks.	Score of map components: number of concepts and propositions, the hier-archy and the branching, and quality of the map based on overall impression.
Fisher, 1990	Task 1. Enter concepts and relation names in the computer with as many links as desired. Task 2. Fill-in-the-blank when a central concept is masked and the other nodes are provided.	Computer response in both tasks. Students construct their maps on a blank screen for Task 1, and filled in the node(s) in a skeleton map for Task 2.	The author only proposed the SemNet computer program as an assessment tool, but did not present any scoring system to evaluate the maps.
Lomask, Baron, Greig, & Harrison, 1992	Write an essay on two central topics on biology (i.e., growing plant and blood transfusion).	Paper-and-pencil response. Trained teachers construct a map from students' written essay. No effort was made to elicit any hierarchy.	Comparison with a criterion map. Two structural dimensions were identified for the comparison: the *size* and the *strength* of structure. The final scored was based on the combination of both dimensions.

Nakhleh & Krajcik, 1991	Semi-structured interview about acids and bases.	Oral response. The interviewer drew three concepts maps—one for acids, one for bases, and one for pH—based on statements that revealed the student's pro-positional knowledge.	Score based on map components: Propositions and examples, cross-links, hierarchy. Experts' maps were used to identify critical nodes and relationships.
Wallace & Mintzes, 1990	Construct a hierarchical concept map from ten given concepts on life zones.	Paper-and-pencil response. Students drew the concept map on a blank page.	Score based on map components: number of relationships, levels of hierarchy, branchings, cross-links, and general-to-specific examples.

adequacy of propositions. Finally, since our focus is on large-scale assessment, we eliminated mapping techniques that required one-to-one interaction between student and tester on practical grounds.

Study 1: Psychometric Properties of Concept Maps

As mentioned previously, concept maps have been used more for instruction than for formal assessment. As a consequence, reliability and validity issues associated with knowledge structure interpretations of concept maps have largely been ignored. Here we highlight several important psychometric issues in the interpretation of concept maps.

Reliability

Reliability refers to the consistency of scores assigned to students' concept maps. Some reliability questions that should be raised about concept maps are: Can raters reliably score concept maps? Are students' scores stable across short occasions? Are scores sensitive to concept sampling from a domain? Reliabilities reported in the literature suggested high interrater reliability and agreement (e.g., Barenholz & Tamir, 1992; Lay-Dopyera & Beyerbach, 1983; Lomask et al., 1992; Nakhleh & Krajcik, 1991). However, these findings should be interpreted cautiously. In some studies the scoring system involved only counting the number of certain map components (e.g., number of nodes; Lay-Dopyera & Beyerbach, 1983). In others, raters scored only a small sample of concept maps (e.g., Nakhleh & Krajcik, 1991). It may be that reliability depends on the scoring criteria (e.g., validity of the propositions) and the number of concept maps used.

We found only one study that reported retest reliability (stability) of concepts map scores. Lay-Dopyera and Beyerbach (1983) concluded that their study failed to establish concept maps as stable measures (stability coefficients ranged from .21 to .73 for different map component scores).

No study has examined the issue of concept-sampling variability. Hence one purpose of Study 1 was to do so. We randomly sampled concepts from a subject domain (Sample A and Sample B) to examine concept sampling variability of map scores.

Validity

Beyond reliability it is essential to justify an interpretation of a concept map as a measure of some aspects of a student's knowledge structure in a science domain—that is, to demonstrate the validity of proposed construct interpretations (i.e., "cognitive structure") of concept maps. One set of construct validity evidence bears on the content domain. In the context of concept maps, content validity involves evidence of content relevance and representativeness of the concepts used for mapping. One important criterion for evaluating content validity, then, is expert judgment of the representativeness of concepts used in the assessment. Another criterion is expert judgment of the accuracy of students' maps within that domain. Only a few studies reported that "experts" judged the terms and maps as consistent with the subject domain (e.g., Anderson & Huang, 1989, Barenholz & Tamir, 1992, Lomask et al., 1992; Nakhleh & Krajcik, 1991).

Other evidence bearing on the construct validity of cognitive structure interpretations of concept maps is correlational (i.e., concurrent validity, and convergent and discriminant validity). Some studies have shown consistent correlations between concept map scores and measures of student achievement (e.g., Anderson & Huang, 1989), whereas others suggested that concept map scores seem to measure a different aspect of achievement than that measured by multiple-choice tests (e.g., McClure & Bell, 1990; Novak, Gowin, & Johansen, 1983).

In Study 1, we examined the issue of concept sampling/representativeness by varying the source of concept samples: students or testers. One mapping technique asks students to provide the concepts in a domain with which to construct the map; the other provides a set of 10 concepts. To summarize, Study 1 addressed the following reliability questions: How reliable are map scores across raters? If concepts are provided by the tester, are map scores sensitive to the sampling of concept terms? We also addressed the following validity questions: Do concept maps provide sensible representations of knowledge in a domain as judged by subject matter experts? Do different mapping techniques provide the same information about a student's knowledge structure? Do different assessment technique scores correlate differently with traditional multiple-choice test scores?

Method

Participants

This study involved two classes of high school chemistry students taught by the same teacher (with four years of teaching experience), a second chemistry teacher (with seven years of teaching experience), and one chemist (expert, with 10 years of experience in research on water quality). All subjects were drawn from the Palo Alto area. The students, the teachers, and the expert were trained to construct concept maps with the same training program.

Of the original 47 students in the two groups, four were dropped from the data set because of incomplete data. Another three were randomly dropped to provide pilot data on which to try out scoring procedures and to equalize cell sizes (see design below). As a result, data were analyzed for 40 students who were assessed on each of three occasions.

Design

The two classes (groups) were randomly assigned to one of two sequences of concept samples: Sequence 1—Sample A first followed by Sample B (class 1), and Sequence 2—Sample B first followed by Sample A (class 2). Each group of students was tested on three occasions: (a) on the first occasion, students were asked to construct a map with no terms provided; (b) on the second occasion, students were asked to construct a map with the first list of concepts; (c) on the third occasion, students were asked to construct a map with the second list of concepts. The 2 x 3 mixed design had one between-subjects factor, sequence of samples of terms, and one within-subjects factor, occasion.

Domain and Material

The topic “Reactions and Interactions” was selected from the knowledge domain defined by the notion of “big ideas” in physical science contained in the Science Framework for California Public Schools (California Department of Education, 1990). Reactions and interactions involve the study of chemical reactions. This big idea focuses on two issues: What happens when substances change? What controls how substances change? (p. 49). At the high school level, these issues involve, among other topics, understanding atomic structure and the nature of ions, molecules, and compounds. The latter was the topic selected for this study.

The concepts of ions, molecules, and compounds were addressed in the unit “Chemical Names and Formulas” in the chemistry curriculum of the high school where the study was carried out. As with the other units in the curriculum, this unit was taught from the widely used text *Chemistry* (Wilbraham, Staley, Simpson, & Matta, 1990). The chapter “Chemical Names and Formulas” defined the domain for sampling concepts to be used in the study.

We compiled a list of 20 key/core concepts in two ways by (a) asking the chemistry teachers to provide the concepts they thought were most important in the unit, and (b) reviewing the textbook used in class ourselves. The process followed in sampling concepts is described in Appendix A.

Two lists of concepts were created from the 20 concepts. Both lists contained four control concepts in common (i.e., ions, molecules, compounds, and electrons). Two samples of six concepts each were randomly selected from the other 16 concepts on the list to form the two sets of concepts (see Appendix B).

Instrumentation

Here we describe the concept mapping techniques (tasks, response format, and scoring system) and the multiple-choice test of achievement.

Concept map task. The two mapping techniques explored in Study 1 varied the task constraints imposed on students: provision or not of the concepts used in the task. Mapping Technique 1—student provides the concepts—asked students to construct a 10-concept map about "Ions, Molecules, and Compounds." Using the three concepts provided by the topic (i.e., ions, molecules, and compounds) students were asked to select another seven concepts that they thought were important in explaining ions, molecules, and compounds, and construct the map (see Appendix C). Mapping Technique 2—tester provides concepts—asked students to construct a concept map using 10 concepts provided on the instruction page (see Appendix C). In both mapping techniques, students were asked to organize the concepts in any way they wanted; no particular (e.g., hierarchical) structure was imposed. Also, students were encouraged to use as many words as they wanted to label the line between two concepts.

Scoring system. The scoring system focused on two aspects of the students' concept maps: (a) the map components, more specifically, the propositions and the nodes; and (b) the disciplinary validity of the map.

The scoring system was based on a *criterion map* developed by the researchers using the 20 key/core concepts. The goal in constructing the criterion map was to identify those propositions (nodes and links) considered to be "substantial" to the domain, and that students should know at that point in the chemistry course (chapter on "Chemical Names and Formulas").

Based on the 20 key/core concepts, a square-matrix was constructed to define all possible links between pairs of concepts. The entries in a cell of the matrix denoted the relation between a specific pair of concepts. Up to 190 links can be drawn between pairs of 20 concepts (see Appendix D). To determine the "substantial" links, teachers, the expert, and the researchers constructed concept maps. The teachers and the expert constructed their maps based on the concepts they considered important in the chapter. We used the 20 key/core concepts. The concepts selected by the expert are presented in Appendix A and were very similar to those in the key/concept list. By comparing the concepts selected as key/core concepts across the three different sources, we concluded that the concept list was discipline valid.

Teachers' concept maps were expected to provide a benchmark for the "substantial" links students were expected to have after studying the chapter and participating in class. The expert's concept map provided the "substantial" links based on the structure of the discipline. Finally, we constructed a third map that was thought to reflect the "substantial" links in the textbook chapter.

An analysis of the four maps identified 48 "substantial" links. About a third of these links were the same across the four maps. The rest of the links (some found in the expert's map, others in the teachers' maps, others in our map) were carefully analyzed and it was concluded that all of them could be expected from the students and justified as appropriate based on the instructional unit. These 48 links were used to construct a *criterion map* with the 20 key/core concepts. These 48 propositions were considered as "*mandatory*"*:* students should reasonably be expected to provide any one of these propositions at that point in their instruction.

For each link in the criterion map a proposition was developed. The labeled links between pairs of concepts provided by the teachers, the expert, and the researchers varied in the quality of their explication of the relationship. For example, the propositions used by the expert more completely explained the links between concept pairs than those used by the teachers. In fact, the propositions found in the expert's map and the researchers' map were more complete and accurate than those found in the teachers' maps. Furthermore, when students' maps were collected, we found that some students provided more accurate propositions than the ones provided by the teachers.

To account for the variation in the quality of the propositions, we developed a *Proposition Inventory*. This inventory compiled the propositions (nodes and direction of links) provided by the teachers' maps, expert's map, students' maps, and researchers' map and classified each proposition into one of five categories: Valid Excellent, Valid Good, Valid Poor, "Don't Care," and Invalid. Table 3 presents the definition of each category (see Appendix E). For example, the valid excellent proposition between *acids* and *compounds* should be read, according to the direction of the arrow (<), as follows: *compounds* that give off H^+ when dissolved in water are *acids*.

Table 3

Quality of the Propositions

Quality of proposition	Definition
Excellent:	Outstanding proposition. Complete and correct. It shows a deep understanding of the relation between the two concepts. *acids-compounds: < that gives off* H^+ *when dissolved in water are*
Good:	Complete and correct proposition. It shows a good understanding of the relation between the two concepts. *acids-compounds: > are examples of*
Poor:	Incomplete but correct proposition. It shows partial understanding of the relation between the two concepts. *acids-compounds: < form*
Don't Care:	Although valid, the proposition does not show understanding between the two concepts. *acids-compounds: > is a different concept than*
Invalid	Incorrect proposition. *acids-compound: > made of*

The Proposition Inventory provided not only propositions that were considered "mandatory," but also propositions for the "other" possible relations between the pairs of concepts in the Key/Core Concept List. These other propositions were considered as "*possible propositions.*" In this form, any other proposition not contained in the criterion map could also be scored, and credit was given if the proposition was valid.

The Proposition Inventory was judged by the expert and a science educator to determine whether the classification of the propositions was accurate. Both agreed on the classification of the propositions.

The scoring system, based on the criterion map and the Proposition Inventory, evaluated two aspects of the students' concept maps as follows:

1. *Map Components*. Two components of the map were contemplated in the scoring: the propositions and the nodes. The validity of each proposition in a student's map was assessed on a 5-level scale (from 0 for invalid to 4 for valid excellent) according to the classification provided in the Proposition Inventory. The concepts used on the nodes were noted, counted, and classified as

contained/not contained in our list of 20 key/core concepts. This last aspect was especially important for the maps constructed with Mapping Technique 1: student provides the concepts.

2. *Map Discipline Validity*. Three map scores were obtained: (a) a total *proposition validity* score—the total sum of the scores obtained across all propositions; (b) *congruence* score (i.e., the degree to which the student's map and the criterion map converge)—the proportion of valid propositions in the student's map out of all possible mandatory propositions in the criterion map; (c) *salience* score—the proportion of valid propositions out of all the propositions in the student's map.

Three scoring forms were designed, one for each condition: No Concepts, Sample A Concepts, and Sample B Concepts. Appendix E shows the scoring form used to score the concept maps when Sample A was provided to students.

Multiple-choice test. Prior to administering the concept maps, all students received a 15-item multiple-choice test on "Chemical Names and Formulas" designed by the researchers and reviewed by both teachers. The internal consistency reliability of the test was .67.

Training

A training miniprogram was designed to teach students, teachers, and the expert to construct concept maps. The program was piloted with another two groups of high school chemistry students, and minor modifications were made.

The training program was delivered by the same researcher to both groups of students to minimize variability. The training lasted about 50 minutes and had four major parts. The first part focused on introducing concept maps: what they are, what they are used for, what their components are (i.e., nodes, links, linking words, propositions), and examples (outside the domain to be mapped) of hierarchical and nonhierarchical maps. The second part emphasized the construction of concept maps. Four aspects of mapping were highlighted in this part of the program: identifying a relationship between a pair of concepts, creating a proposition, recognizing good maps, and redrawing a map. Students were then given two lists of common concepts to "collectively construct" a map. The first list focused on the theme "water cycle"—a nonhierarchical map; the second list focused on the theme "living things"—a hierarchical map. The third part of the program provided each individual with nine concepts on the theme "food web" to construct a map individually. The fourth and final part of the program was a discussion of students' questions after they had constructed their individual maps.

After students in both classes had been trained, a random sample of 10 of the individually constructed maps was analyzed for each group (a total of 20 concept maps) to evaluate the training. This analysis focused on three aspects of the maps: use of the concepts provided on the list, use of labeled links, and the validity of the propositions. Results indicated that (a) 97.8% of the students used all the concepts provided on the list, (b) 100% used labeled lines, and (c) 93.8% of the propositions provided were valid. We concluded that the program succeeded in training the students to construct concept maps.

Procedure

Study 1 was conducted in three 55-minute sessions during a three-week period. Both classes were assessed on the same days in their respective classrooms. The first session was training. Two weeks of instruction followed. The second and third sessions were conducted consecutively after instruction.

At the second session, students took the multiple-choice test (13 minutes, on average). After all the students finished this test, a 15-minute reminder about concept maps was conducted. Then, students constructed concept maps under the no-concept-provided condition. Although students had about 30 minutes to construct their maps, over 90% of the students finished in 20 minutes.

At the third session, students constructed maps with concepts provided. Class 1 first mapped with Sample A concepts, whereas Group 2 mapped first with Sample B. After students finished their first maps, they constructed the next map using the other sample of concepts. Construction of one concept-provided map took 14 minutes, on average, for both groups.

Results and Discussion

This study addressed three questions: Can two concept mapping techniques be interpreted in the same way, as representing the same aspect of a student's knowledge structure? Do concept maps provide reliable scores. Are cognitive-structure interpretations of map scores valid?

Before turning to these questions a preliminary methodological issue needs to be addressed: Does concept sample sequence (Sample A and Sample B) affect map scores? A 2x3 (sequence by map sample) split-plot ANOVA revealed no significant differences (a = .05) for sequence (S), concept sample (CS), or their interaction (SxCS; F_S = .64, F_{CS} = .33, F_{SxCS} = .18). Since no significant differences were found, we collapsed the two groups and present results overall.

Comparison of mapping techniques. The two mapping techniques were compared as follows: (a) the *"What"* in the students' maps question: What concepts do students recall and use in Mapping Technique 1 (no concepts provided)? Were the same concepts used with technique 1 as those selected in the "key/core concept list" from which Sample A and B were created? If not, were the concepts related in some way to the knowledge domain assessed? (b) the *"How Valid"* from a disciplinary point of view question: Are the "proposition validity" and "salience" scores higher for maps created with concepts selected by the students than for maps constructed with concepts provided?

Map components. Table 4 presents the mean number of key, "other," and total concepts used in the maps along with standard deviations. We focus on concepts students provided under Mapping Technique 1 and examine the characteristics of these concepts (e.g., How relevant are the concepts to the topic assessed? Were the concepts selected by the student the same as those selected in the key/core concept list?). We expected that students with greater competency in the subject matter would recall more key/core concepts than students with less competency.

Table 4

Means and Standard Deviations of Concepts Considered as Key-Core Concepts, "Other" Concepts, and the Total Number of Concepts Used by Students Across the Three Conditions

Mapping conditions	Key/core concepts		Other concepts		Total number	
	Mean	*SD*	Mean	*SD*	Mean	*SD*

No Concepts	6.40	(2.05)	3.28	(2.54)	9.68	(2.20)
Sample A	9.88	(.33)	.10	(.50)	9.98	(.62)
Sample B	9.83	(.45)	.03	(.16)	9.85	(.43)

With technique 1, about 75% of the students used six or more key/core concepts that were also found in our list. Only one student used 10 concepts that were all key/core concepts. Besides ions, molecules, and compounds (the three concepts provided in the instructions; see Appendix C), students used anions, cations, and acids most frequently for their maps. From the other 16 concepts on the list, 14 (e.g., binary ionic compounds, polyatomic ions, bases) were used by at least one student. Binary molecular compounds and neutral charge were not used by any student. The last finding was a surprise because the knowledge that compounds have a neutral charge is a key idea in the topic assessed.

The correlation between the number of key/core concepts on the students' concept maps with the proposition validity total score was moderate and significant (r = .56, p = < .05). Those students who recalled more key/core concepts tended also to have a greater valid propositions total score.

The “other” concepts provided by the students were classified as “related,” but not important, and “not related” (e.g., plants, animals) to the knowledge domain assessed. About 63% of the “other” concepts were considered *related* to the topic “Ions, Molecules and Compounds” (e.g., element, ductile, HCl). From all these concepts, “element” was the most frequently used as an “other” concept. Forty-seven percent of the students used it for their maps.

It is important to mention that “element” was on the original 23-concept list; however, we decided to drop it for two reasons: Teachers did not select it, and only a few links with the others concepts on the list could be drawn. This was a big “miss” on the key/core concepts list.

Another missing set of items on this list were “examples of compounds.” Examples of molecular compounds, polyatomic ions, acids, or ternary ionic compounds could have indicated whether students understood, say, how HCl is related to anions, cations, and binary ionic compounds, instead of only memorizing the formula (i.e., does the student understand that HCl is an acid and is a binary ionic compound that has one simple anion, Cl^-, and a cation, H^+?).

Unrelated concepts included, for the most part, “general” chemistry concepts (e.g., symbols, mixtures, substances). Even though they can be related in some way to the other key/core concepts, they do not reflect students’ understanding about ions, molecules, and compounds. The correlation between the number of “other” concepts and the proposition validity total score was close to zero ($r = -.10$, $p = > .05$). It is important to remember that students were given credit for any valid proposition in their maps, related or not, to the topic at hand.

Not surprisingly, it seems that if a student has an adequate understanding of the topic he/she will provide topic relevant-concepts and the propositions between pairs of concepts will tend to be valid. When a student’s understanding is not very good, he or she will provide more nonrelevant “other” concepts, which result in superficial (i.e., “don’t care” type of propositions) and/or invalid propositions.

Little can be said about the “What” question for Samples A and B because the concepts were provided. An average of about 86% of the students over the two conditions used the 10 concepts provided in the instructions. Only one student in Sample B used only eight of the concepts. In Sample A, a few students used 13 concepts to construct the map, 10 provided on the list and three “other” concepts.

Concept map validity. The disciplinary validity of students’ concept maps was evaluated using the following data: (a) the proposition validity score, (b) the proportion of valid propositions in a student’s map to the number of “mandatory” propositions (*congruence*), and (c) the proportion of valid propositions to all the propositions in that student’s map (*salience*). Table 5 shows the means and standard deviations for proposition validity, congruence and salience scores across the three conditions.

Students’ knowledge about “Ions, Molecules, and Compounds” was partial and not close to the criteria established by the criterion map. The low proposition validity and congruence scores across the three conditions indicate that students’ knowledge was rather weak compared to the criterion map.

Salience mean score shows that about half of the propositions provided by the students were valid.

Table 5

Means and Standard Deviations of the Proposition Validity, Congruence, and Salience on Each Condition by Each Group

	No Concepts		Sample A		Sample B	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Proposition validity	11.98[a]	(7.62)	11.76	(8.86)	12.52	(7.20)
	(Max=180)		(Max=108)		(Max=124)	
Congruence	b		.17	(.10)	.16	(.08)
			(Max. # of Prop=27)		(Max. # of Prop=31)	
Salience	.51	(.25)	.47	(.27)	.51	(.24)

[a] Maximum score was calculated based on 45 excellent valid propositions students could provide if 10 concepts were used in a map.

[b] Proportions were not calculated since no criterion could be established to determine the expected number of propositions.

In the no-concept-provided condition students had the opportunity to select concepts they felt most confident about to reflect their knowledge of the topic, but still only half of their propositions were valid. Since credit was given to all valid propositions provided by students in their maps under this condition, points were credited for knowledge that was not essential to the topic “Ions, Molecules, and Compounds.” For example, students provided “other” concepts, like chemistry or chemical substances, and related them in such a way that the explanation of the relation was valid (e.g., chemical substances are studied in chemistry—a “don’t care” proposition), but do not provide any evidence of students’ knowledge of the topic.

To test mean differences among the three conditions, two Hotelling’s T^2 tests for repeated measures were carried out, one for the proposition validity scores and the other for the salience scores. Differences in congruence means were not tested because this score could not be calculated for the No Concepts condition (see note b in Table 5).

Although we report regular salience score means, the analysis of the salience scores was carried out using the natural log transformation of the proportions. No significant differences were found between the proposition validity means (T^2 = .9809; p = > .05) or the salience means (T^2 = 2.4367; p = > .05). The two concept mapping techniques—students provided concepts or tester provided concepts—yielded similar mean map scores.

To evaluate whether different samples of concepts, Sample A and Sample B, influenced students' map scores, a series of Hotelling's T^2 tests were carried out between Sample A and Sample B, one for each score obtained (i.e., proposition validity, congruence, and salience scores). To adjust the number of expected propositions (links) in Sample A and Sample B we transformed the proposition validity total score into a proportion (i.e., student's total score divided by the total score of all "mandatory" propositions in the criterion map, this is 108). Although raw score means are presented in the table, all statistical analyses were carried out using a natural log transformation of the proportions.

No significant differences (a = .05) were found on any set of scores (proposition validity: T^2 = 1.2642; congruence: T^2 = .7821; salience: T^2 = 3.4990). Students' map scores did not differ, on average, in Sample A and Sample B. Probably the procedure used in selecting the concepts (see Appendix A) helped to create a list of cohesive concepts; therefore, any combination of concepts could provide critical information about a student's knowledge about the topic.

Reliability and Validity of Concept Maps

We examined the generalizability of proposition validity scores across raters and concept samples in the context of G theory (Table 6). The first G study considered the three conditions, No Concept, Sample A and Sample B; the second one considered only Sample A and Sample B. In general, raters did not introduce error variability into the scores (percent of total variability is negligible).

In both G studies, the largest variance component was for persons followed by the interaction of person x condition (or sample in the second part of the table). Not surprisingly, students' relative standing varied from one condition (or sample of concepts) to the next (some students did better in Sample A, others with Sample B, and still others when they selected the concepts to use in their maps). These results are consistent with what has been found with science performance assessments. The interaction of person x task has been a major source of unreliability (Shavelson, Baxter, & Pine, 1991).

Table 6

Estimated Variance Components and Generalizability Coefficients for a Person x Rater x Condition G Study Design for No Concept, Sample A, and Sample B Conditions Using the Proposition Validity Score

Source of variation	Estimated variance components	Percent of total variability
Three conditions (NC, A, B)		
Persons (p)	46.43333	71.64
Rater (r)	0.09712	.15
Condition (c)	0*	0
p x r	0*	0
p x c	14.78953	22.819
r x c	.00951	.01
prc,e	3.48216	5.372
ρ^2 ($n_r = 2$, $n_c = 3$)	.89	
ϕ	.89	
Sample A and Sample B		
Persons (p)	52.33301	78.67
Rater (r)	0*	0
Sample (s)	0*	0
p x r	.52564	.79
p x s	11.73782	17.64
r x s	.12532	.188
prs,e	1.79343	2.696
ρ^2 ($n_r = 2$, $n_s = 2$)	.88	
ϕ	.88	

* Negative variance components set to zero; in no case was the variance component more than -0.26.

Both relative and absolute generalizability coefficients are high, suggesting that concept map scores can consistently rank students relative to one another ($\hat{\rho}^2$) as well as provide a good estimate of a student's level of performance, independently of how well classmates performed (ϕ).

Another set of G studies was carried out for congruence and salience scores. Patterns of variability were the same across the two different sets of scores (i.e., the highest percentage of variability was for persons followed by the interaction person x condition). Relative and absolute coefficients were roughly of the same magnitude for both types of scores although they were lower than those found when proposition validity scores were used (Congruence: $\rho^2 = .80$, $\phi = .80$; Salience $_{(NC, A, B)}$: $\rho^2 = .79$, $\phi = .79$; Salience $_{(A,B)}$: $\rho^2 = .81$, $\phi = .81$).

Two decision (D) studies were carried out (variance components on the second part of Table 6 were used in the computation) to determine the magnitude of the coefficients by varying the number of conditions of the two facets: (a) when two raters and only one sample of concepts are used, and (b) when one rater and one sample are used. These conditions represent a more realistic situation in large-scale assessment. Although both relative and absolute coefficients were lower in magnitude (Two raters, one sample of concepts: $\rho^2 = .80$, $\phi = .80$; One rater, one sample of concepts: $\rho^2 = .78$, $\phi = .78$), still both are reasonably high.

Finally, if concept maps measure somewhat different aspects of declarative knowledge than multiple-choice tests, the correlation between these two measures should be positive, because they measure the same knowledge domain, but moderate in magnitude. The correlations in Table 7 are consistent with these expectations. All are positive and moderately high. We interpret these findings to mean that concept maps and multiple-choice tests measure overlapping and yet somewhat different aspects of declarative knowledge.

Table 7

Correlation Between the Multiple-Choice Test and Proposition Validity, Congruence, and Salience Score

	No Concepts	Sample A	Sample B
Proposition validity	.58	.64	.63
Congruence	a	.66	.55
Salience	.45	.61	.50

[a] Not calculated (see note b on Table 5).

A Closer Look at the Students' Maps

Both teachers constructed a concept map on the topic "Ions, Molecules, and Compounds." When reviewing the teachers' maps, we identified a misconception in the map of the teacher who taught the students who participated in this study. According to the teacher's map, anions and cations lose or gain electrons when, in fact, atoms lose or gain electrons to become ions; and ions, according to their charge—positive or negative—are either cations or anions. We decided to take a closer look at the students' maps to find out whether this misconception was reproduced by the students in their maps. About 17% of the students showed exactly the same misconception.

Another phenomenon observed, but not evident by simply examining the scores, is how the sample of concepts seems to "lead" some students to create connections that, probably due to their partial knowledge, result in invalid propositions. We observed that more students related molecules to ions with Sample B than with Sample A concepts. Sample B's limited selection of concepts that could be related to molecules "forced" students to look for more connections with the concept, even though these connections were wrongly conceived.

Conclusions

Study 1 explored the potential of two concept mapping techniques for use in large-scale science assessment. We examined (a) whether map scores were sensitive to who chooses the concepts to be used in the map (student or tester) and to the sampling of the concepts (e.g., random samples of key concepts from a domain), and (b) how reliable and valid concepts map scores are.

Our findings lead to the following tentative conclusions:

1. The two mapping techniques explored seem to provide equivalent interpretations about students' knowledge structure. Both techniques provide similar students' scores. However, note that under the No Concept condition, students were probably given "too much" credit for propositions that did not provide evidence about their knowledge of the topic assessed. We plan to further explore this issue before reaching a final conclusion about the equivalence of these mapping techniques.

2. Randomly sampling concepts provides equivalent map scores, at least when a careful procedure has been followed in selecting the set of key/core concepts.

3. Concept maps can be reliably scored, even when judgment about the quality of the propositions enter into the scores.

4. Students' concept map scores appear to generalize across samples of key/core concepts.

5. The relationship between multiple-choice test and concept maps suggests that they measure overlapping and yet somewhat different aspects of declarative knowledge.

Moreover, we have found that students can be trained to construct concept maps in a short period of time with limited practice. This, from the large-scale assessment perspective, is important, not only because training time may no longer be an issue, but also because students' facility in using concept maps has been demonstrated. From a closer look at students' maps we know that even though practice may improve map characteristics, this training shows that students were able to demonstrate their knowledge in the topic assessed.

It appears that there is potential for using concept maps in an assessment of science achievement. Still more questions need to be answered before we can conclude that concept maps reliably and validly evaluate students' knowledge structure.

Study 2: Hierarchical Structures in Concept Maps

A common practice when using concept maps is to ask students to construct hierarchical maps (e.g., Novak et al., 1983; Markham, Mintzes, & Jones, 1994; Roth & Roychoudhury, 1993). No attention has been directed to how the instructions interact with the structure of the subject domain to be mapped. This interaction is the focus of Study 2.

Methodologically and conceptually, there is no need to impose a hierarchical structure if the structure of the content domain to be mapped is not hierarchical. In fact, it may be that different map structures are needed to represent different types of content structures. For example, Harnisch, Sato, Zheng, Yamaji, and Connell (in press) proposed the use of "chain maps" to represent procedural or sequential activities. Regardless of the type of organization, we expect that as subject matter mastery increases, the structure of the map should increasingly reflect the structure, hierarchical or not, in the domain as held by experts.

For identifying the structure of a domain, we need to assume that there is some “ideal organization” that best reflects the structure, and that “experts” in that domain possess that ideal organization to some degree. Experts’ knowledge structures are assumed to be highly connected and articulated (e.g., Glaser, in press). But, do all experts in a field share the same knowledge structure? Acton, Johnson, and Goldsmith (1994) showed that experts’ structures are highly variable. Indeed, individual differences in experts’ maps will arise because knowledge structure should reflect not only domain knowledge, but also a personal schema for thinking and cognitive activity (e.g., strategies for problem solving and interpretation; Glaser, in press). Therefore, we expected different experts to provide somewhat different concept maps; and consequently, inferences about the structure of a subject domain from one expert’s knowledge structure to another might also vary.[3]

Assuming that any expert’s knowledge structure provides an accurate representation of the subject domain, how can we determine whether the structure is hierarchical? The identification of hierarchical structures from the natural (i.e., inorganic and organic), conceptual, and artifactual worlds (e.g., computer language, social events) has been a topic of discussion for the last three decades (e.g., Dress & von Haeseler, 1990; Whyte, Wilson, & Wilson, 1969). Unfortunately, the term *hierarchy* has been considered a “catch-all” term used to cover a variety of related yet distinct notions (e.g., Bunge, 1969; Green, 1969; Mesarovic & Macko, 1969). This makes it difficult to find a formal definition that can be used without controversial results (e.g., Dress & von Haeseler, 1990; Green, 1969; Jones, 1969; Rosen, 1969).

Bunge, in 1969, proposed a formal definition of hierarchy: “Strictly speaking, a hierarchy or hierarchical structure is a set equipped with a relation of domination or its converse, subordination” (p. 17). According to his definition, H (i.e., a set of elements with binary relations) is a hierarchy *if and only if*: (1) H has one and only one beginner element—”a supreme commander”; (2) no matter how low in the hierarchy an element is, it is under

---

[3] An entire study could be carried out on similarity of knowledge structures among experts.

the command of the beginner; (3) every member has a single boss; (4) the relation among the elements is antisymmetric and transitive (in Bunge's colloquial terms, "Togetherness but no back talking," p. 16); and (5) the relation between elements is a relation of domination or power (i.e., elements are held together by a subordinate relation). According to Bunge, any structure has to meet each of the five assumptions if it is to qualify as a hierarchy. In sum, "a diagram of a hierarchy is a finite tree branching out of a single point (namely *b*) and no loops" (Bunge, 1969, p. 19).[4]

According to this definition "pure" hierarchical concept map structures may be difficult to find: maps constructed by experts or knowledgeable students may not comply with criteria 2, 3, 4 and 5 since highly connected structures with crosslinks across levels and between branches are typical of mastered knowledge. Therefore, "degree of hierarchiness," may be a more accurate way to describe concept map structures. A concept map that has more than one beginner node, many nodes with more than "one boss," many cycles (or loops), and concepts that are not held together by subordinate relations exclusively, can be considered "less hierarchical than" a map that has one beginner node, no nodes with more than "one boss," no cycles, and concepts that are held together primarily by subordinate relations.

Defining the structure of a particular content domain, then, is not an easy task. Different conclusions about the structure may arise if different experts and criteria are used.

In this study, we examined the intersection of the task demands and constraints with the structure of the subject domain to be mapped. Two mapping techniques with the same task demand (i.e., construct a map) but different task constraints (i.e., imposing on students a specific structure for their maps) were used. Mapping Technique 1 asked students to construct their concept maps using a hierarchical structure, and Mapping Technique 2 asked students to construct their maps organizing the concepts in any way they

---

[4] If a structure is hierarchical, another characteristic emerges: the *levels*. Level is an ambiguous term that is also the object of philosophical debate (e.g., Bunge, 1969; Mesarovic & Macko, 1969). A level can be considered as an "assembly of things of a defined kind, e.g., collection of systems characterized by a definite set of properties and laws . . ." (Bunge, 1969, p. 20). To define a hierarchy level, then, we should, for example, evaluate whether every member at a certain level shares an exclusive property that makes that level different from another level.

wanted. To evaluate the intersection of imposing a structure with the structure of the subject domain, two content domains were selected, one with a "hierarchical" structure and another one with a "nonhierarchical" structure as held by two experts. If the structure of a map should reflect the structure in the domain as held by an expert, we expected that students who knew the subject matter would construct maps with similar structures to that of the expert.

Method

Participants

Two classes of high school chemistry students taught by the same teacher (with seven years of teaching experience), a second chemistry teacher (with five years of teaching experience), and two experts, one chemist (with 10 years of experience in research on water quality) and a physicist (with 14 years of experience in research on subatomic particles), participated in Study 2. As in Study 1, the students, the teachers and the experts were trained to construct concept maps with the same training program. All subjects were drawn from the Palo Alto area.

Of the original 62 students in the two groups, eight students were dropped from the data set because of incomplete data. Another six students were randomly dropped to provide pilot data to check out scoring procedures and equalize cell sizes. As a result, data for 48 students were analyzed.

Design

Two topics were selected as having different structures according to the criterion maps: one topic with a hierarchical content structure and another one with a nonhierarchical content structure. Classes were randomly assigned to the topic in which they were assessed. Within each class, students were randomly assigned to one of two mapping techniques: Mapping Technique 1—Instructions imposing the construction of hierarchical maps (Hierarchical Instructions); and Mapping Technique 2—Instructions without restrictions on the type of structure for constructing their maps (Nonhierarchical Instructions). This factorial design had two between-subjects factors: (a) Topic, with two levels: topic with hierarchical structure and topic with nonhierarchical structure; and (b) Mapping Technique, with two levels: Hierarchical Instructions and Nonhierarchical Instructions.

Domain and Material

The two topics selected for this study were “Atomic Structure” and “Nature of Ions, Molecules, and Compounds,” which are topics involved in the big idea “Reactions and Interactions” as described in the *Science Framework for California Public Schools* (California Department of Education, 1990). These two topics were taught as two consecutive units in the chemistry curriculum at the school where the study was conducted.

Both units were taught using the chapters “Atom Structure” and “Chemical Names and Formulas” of the textbook *Chemistry*, used by Study 1 students (Wilbraham, Staley, Simpson, & Matta, 1990).

Two experts were used to define the content structure of the two topics. According to their area of expertise, the two experts were asked to construct a concept map on either “Atom Structure” or “Ions, Molecules, and Compounds.” The “hierarchiness” of the experts’ maps were judge based on four aspects: (a) the number of “beginner” nodes (i.e., nodes with only arrows coming out but no arrows coming in); (b) the number of nodes with more than “one boss” (i.e., nodes with more than one arrow coming into the node); (c) the number of cycles or “loops” in the map; and (d) the percentage of subordinate propositions. The expert’s map on the topic “Atom Structure” had one “beginner” node; three nodes with more than one arrow coming in; no cycles; and 95% of the propositions in his map were subordinate. The expert’s map for “Ions, Molecules, and Compounds” had three “beginner” nodes; 10 nodes with more than one arrow coming in; no cycles; and 97% of the propositions in his map were subordinate. Based on this information, the topic “Atom Structure” was considered as having a more hierarchical structure than the topic “Ions, Molecules, and Compounds.”

The textbook chapters “Atom Structure” and “Chemical Names and Formulas” were used to defined the domain for selecting the concepts used in the study. A list of 17 key/core concepts (see Appendix F) was compiled from the “Atom Structure” chapter using the same procedure described in Study 1 (see Appendix A). For the chapter “Chemical Names and Formulas” (the same topic used in Study 1), we eliminated from the 20 key/core concept list the three concepts (i.e., binary molecular compounds, negative charge, and positive charge) that had the least number of connections with other concepts based on the criterion map.

Instrumentation

This section describes the concept mapping techniques (tasks, response format, and scoring system) as well as the multiple-choice test of achievement.

Concept map task. The two mapping techniques explored in this study varied in the task constraints imposed on the students: constructing a hierarchical or nonhierarchical map. Mapping Technique 1—hierarchical structure imposed—asked students to construct a 17-concept map organizing the more general terms above more specific terms (see Appendix G). Mapping Technique 2—no specific structure imposed—asked students to construct a 17-concept map organizing the terms in any way they wanted (see Appendix G).

Scoring system. As in Study 1 the scoring system was based on a *criterion map*—a composite of the experts', teachers', and researchers' maps. Two 17-concept criterion maps were constructed to identify those propositions "substantial" to the domain and that students should know about "Atom Structure" and "Ions, Molecules, and Compounds" at that point in the chemistry course.

Up to 136 links can be drawn between the pairs of the 17 concepts. The analysis of the maps constructed by the teachers, the expert, and the researchers identified 25 "mandatory" propositions for the topic "Atom Structure," and 44 "mandatory" propositions for the topic "Ions, Molecules, and Compounds."

To score the validity of the students' map propositions we used two proposition inventories, one for each topic. The Proposition Inventory constructed for Study 1 was used to score the "Ions, Molecules, and Compounds" concept maps. The Proposition Inventory constructed for "Atom Structure" also compiled and classified the propositions provided in the teacher's map, the expert's map, the researchers' map and the students' maps into the categories described in Table 3 (i.e., Valid Excellent, Valid Good, Valid Poor, "Don't Care," and Invalid). The "Atomic Structure" Proposition

Inventory was judged by the expert (i.e., the physicist) to determine the validity of the classification of the propositions. No changes were necessary.

Both inventories included the "mandatory" and the "possible" propositions (i.e., the 136 possible links between the pairs of the 17 key/core concepts). Therefore, students were credited for any valid proposition that they provided that was not contained on the criterion map.

To score the "hierarchiness" of the map structures, we evaluated four aspects of the student's map structure as to whether (a) the map had only *one* "beginner or commander" node (i.e., a node that had only arrows coming out, but *none* coming in), (b) the nodes had a "single boss" (i.e., on each node only *one* arrow comes in), (c) the relations among nodes were antisymmetrical and transitive (i.e., no cycles in the structure), and (d) the relations between pairs of nodes were subordinate (i.e., for each proposition, one of the nodes is considered to be less general, or have an inferior rank, or be under the control of the other). Information provided by these four aspects bears, directly or indirectly, on the five criteria proposed by Bunge to classify a structure as hierarchical.

The scoring system evaluated three aspects of the students' concept maps as follows:

1. *Map Components*. Three components of the map were considered in the scoring: the propositions, the nodes, and the labeled links (i.e., the arrows). As in Study 1, the validity of each proposition in a student's map was assessed on a 5-level scale (from 0 for invalid to 4 for valid excellent) according to the classification provided in the Proposition Inventories. Each *valid* proposition in the student's map was also judged as having a subordinate relation (1) or not (0). The concepts used on the nodes were noted and counted. Nodes that had only arrows coming out but not coming in were noted and counted as well as the nodes that had more than one arrow coming in. The path of the links (arrows) was analyzed to identify cycles (i.e., loops: direct or indirect symmetrical and nontransitive relations) in the student's map. The number of cycles was noted.

2. *Map Discipline Validity*. As in Study 1, three map validity scores were obtained: (a) a total *proposition validity* score—the total sum of the scores obtained across propositions; (b) a *congruence* score (i.e., the degree to which

the student's map and the criterion map converge)—the proportion of valid propositions in the student's map out of all possible mandatory propositions in the criterion map; (c) a *salience* score—the proportion of valid propositions out of all the propositions in the student's map.

3. *Map Hierarchiness*. Four map "hierarchiness" scores were obtained: (a) *beginner-nodes* score—total number of nodes with arrows out going but no incoming arrows in the student's map; (b) *cycle* score—the total sum of cycles observed on each student's map; (c) *circuit* score—the total sum of nodes with more than one incoming arrow in the student's map; (d) *subordinate* score—the proportion of subordinate propositions in the student's map out of all valid propositions in the student's map.

One form was designed to score the maps for each topic. Appendix H shows as an example the scoring form used to score the "Atom Structure" concept maps.

Multiple-choice test. Prior to administering the concept maps, both classes received a 15-item multiple-choice test: Group 1 received the test on "Atom Structure" and Group 2 on "Ions, Molecules, and Compounds." The multiple-choice tests were designed by the researchers and reviewed by the teachers. The internal consistency reliability was .56 for the "Atom Structure" test and .71 for the "Ions, Molecules, and Compounds" test. Three unrelated items were dropped from the atom structure test to increase the internal consistency coefficient.

Training

The training program for constructing concept maps used in Study 1 was also used in this study to train students, the teacher, and the expert. To evaluate the training, a random sample of 10 of the individually constructed maps was analyzed for each group (a total of 20 concept maps). Results indicated that (a) 100% of the students in Group 1 (those who studied the "Atom Structure" topic) and 97.8% in Group 2 (those who studied the "Ions, Molecules, and Compounds" topic) used all the concepts provided on the list; (b) 100% of the students in both groups used labeled lines; and (c) 85.6% and 89.5% of the students' propositions, in Groups 1 and 2 respectively, were valid. We concluded that the training program succeeded in training the students to construct concept maps.

Procedure

Study 2 was conducted in three 55-minute sessions during a four-week period. The first session was used for training. In the second session, students took the multiple-choice test. In the third session, students received, first, a 15-minute reminder on how to construct concept maps, and then they were asked to construct the concept maps on the topic.

Both classes were trained on the same day in their respective classrooms before the unit "Atom Structure" was taught. Group 1 had the second and third sessions two weeks after the training, when the instruction of the unit "Atom Structure" ended. Group 2 received the sessions four weeks after the training, when the instruction of the unit "Chemical Names and Formulas" ended. Construction of concept maps took 25 minutes, on average, for both groups.

Results and Discussion

Study 2 addressed the questions: Is there an effect of imposing a hierarchical structure (Mapping Technique 1) and nonhierarchical structure (Mapping Technique 2) on students' representations of two types of content domains?

According to the 2 x 2 factorial design, the characteristics of the four groups were: Group 1—students mapped the hierarchical structured topic (i.e., Atom Structure) and received Mapping Technique 1 with instructions imposing a hierarchical structure for constructing their maps (HT/HI); Group 2—students mapped the hierarchical structured topic and received Mapping Technique 2 with instructions not restricting the type of structure for constructing their maps (HT/NHI); Group 3—students mapped the nonhierarchical structured topic (i.e., Ions, Molecules, and Compounds) and received Mapping Technique 1 (NHT/HI); and Group 4—students mapped the nonhierarchical structured topic and received Mapping Technique 2 (NHT/NHI).

Groups were compared as follows: (a) *Validity* of the students' maps from a disciplinary point of view: Do the proposition validity, congruence, and salience scores differ across groups? And (b) *Hierarchiness* of the students' map structures: Do students' map structures differ in the degree of hierarchiness according to the mapping technique and the structure of the topic? (E.g., was the hierarchical structured topic mapped in a hierarchical way even though no instructions to do so were provided?)

Concept map validity. The disciplinary validity of students' maps was evaluated using the following data: (a) the proposition validity score, (b) the proportion of valid propositions in a student's map to the number of "mandatory" propositions (congruence score), and (c) the proportion of valid propositions to all the propositions in a student's map (salience score).

Two raters scored each student map. Interrater reliability coefficients were typically high: .98 for the proposition validity total score, .99 for the congruence score, and .98 for the salience score.

Table 8 shows the means and standard deviations for proposition validity, congruence and salience scores across groups.

Table 8

Means and Standard Deviations of the Proposition Validity, Congruence, and Salience Scores on Each Condition

	Proposition validity		Congruence		Salience	
Group	Mean	*SD*	Mean	*SD*	Mean	*SD*
Atoms						
1 HT/HI	36.92 (Max=100)[a]	(13.66)	.51 (Max. # of Prop=25)	(.16)	.76	(.15)
2 HT/NHI	40.29 (Max=100)	(9.64)	.54 (Max. # of Prop=25)	(.11)	.78	(.15)
Ions, Molecules, and Compounds						
3 NHT/HI	28.29 (Max=176)[b]	(15.73)	.23 (Max. # of Prop=44)	(.12)	.57	(.26)
4 NHT/NHI	28.96 (Max=176)	(15.13)	.23 (Max. # of Prop=44)	(.11)	.60	(.25)

[a] Maximum score was calculated based on 25 excellent valid mandatory propositions students could provide.

[b] Maximum score was calculated based on 44 excellent valid mandatory propositions students could provide.

Mean proposition validity scores across the groups revealed that students' knowledge was partial and not close to the standard established by the criterion maps. Lower mean scores observed for Groups 3 and 4 indicated that students' knowledge about "Ions, Molecules, and Compounds" was weaker than students' knowledge about "Atom Structure" when compared with the criterion maps. The same pattern is observed in the congruence and salience mean scores.

It is important to note that the topic "Ions, Molecules, and Compounds" was more complex than the topic "Atom Structure." The number of mandatory links in the "Ions, Molecules, and Compounds" criterion map almost doubled that of the "Atom Structure" criterion map. This difference in complexity was also reflected in the experts' maps. The "Atom Structure" expert's map had half of the links observed in the "Ions, Molecules, and Compounds" expert's map.

To evaluate the interaction of topic and mapping technique three 2 x 2 factorial ANOVAs were carried out, one for each proposition score. Results for proposition validity score indicated no significant interaction of topic by mapping technique ($F_{TxMT} = .116$; $p > .05$), and, not surprisingly, a significant topic main effect ($F_T = 6.32$; $p < .05$), although this result is not of special interest for our purposes.

ANOVA results for congruence and salience scores also found no significant interaction and a significant topic effect (Congruence: $F_{TxMT} = .024$, $p > .05$; and $F_T = 29.87$; $p < .05$; and Salience: $F_{TxMT} = .338$, $p > .05$; and $F_T = 5.46$; $p < .05$). Although regular means are reported in Table 8, the analyses for the congruence and salience scores were carried out using the natural log transformation of the proportions.

To evaluate the extent to which concept maps measure different aspects of declarative knowledge than multiple-choice tests, the correlation between these two measures was calculated. Coefficients are presented in Table 9. Correlations are positive and vary from moderate to moderately high. We interpret these correlations to mean that both tests measured the same knowledge domain, but still somewhat different aspects of it.

Table 9

Correlation Between the Multiple-Choice Test and Proposition Validity, Congruence, and Salience Scores by Topic

	Proposition validity	Congruence	Salience
Atom Structure	.52	.36	.33
Ions, Molecules, and Compounds	.43	.39	.42

Note. Correction for attenuation was calculated for each coefficient.

Concept map hierarchiness. The "hierarchiness" of the students' maps was evaluated using the following data: the number of beginner-nodes, the number of cycles, the number of circuits, and the proportion of subordinate propositions out of all the propositions in the student's map.

Only the maps of the top 25% of the students in each group were evaluated for "hierarchiness." Those students with low scores did not have sufficient knowledge to reflect a suitable structure in content domain. The mean scores of the four students from each group are presented in Table 10.

Table 10

Means and Standard Deviations of the Proposition Validity, Congruence, and Salience Scores Considering Only the Four Top Students in Each Group

Group	Proposition validity		Congruence		Salience	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
1 HT/HI	51.25	(7.60)	.65	(.09)	.87	(.01)
	(Max=100)[a]		(Max. # of Prop=25)			
2 HT/NHI	49.75	(5.19)	.63	(.07)	.94	(.06)
	(Max=100)		(Max. # of Prop=25)			
3 NHT/HI	44.37	(9.92)	.34	(.05)	.80	(.04)
	(Max=176)[b]		(Max. # of Prop=44)			
4 NHT/NHI	44.87	(10.03)	.32	(.09)	.77	(.10)
	(Max=176)		(Max. # of Prop=44)			

[a] Maximum score was calculated based on 25 excellent valid mandatory propositions students could provide.

[b] Maximum score was calculated based on 44 excellent valid mandatory propositions students could provide.

The pattern of mean scores observed for all students is similar for this selected group of students; means were higher for Groups 1 and 2 on the three map validity scores. However, students' maps still indicated very partial knowledge about the topics when compared with the criterion maps.

The "hierarchiness" of the top students' maps were scored by two raters. Interrater reliability coefficients across hierarchiness scores were also high: .82 for beginner nodes; 1.00 for cycles; .98 for circuits, and .87 for subordination. Table 11 presents the mean for each hierarchiness score.

To evaluate whether an interaction effect—topic by mapping technique—was observed, a factorial ANOVA was carried out for each of the map scores. No significant interaction or main effect was found in any of the "hierarchiness" scores (Beginner Nodes: $F_{TxMT} = .67$, $p > .05$; Circuits: $F_{TxMT} = .009$, $p > .05$; and Subordination: $F_{TxMT} = .34$, $p > .05$). These results indicated that imposing a hierarchical structure does not interact with the structure of the content domain mapped. However, this interpretation seems premature since some problems arose in the way "hierarchical structure" was defined. For example, according to the four "hierarchiness" criteria used, no pure "hierarchical" structures could be identified in any of the student maps. However, the high proportion of subordinate relations suggest a high degree of hierarchiness in all the student maps, independent of the condition. Furthermore, when only subordinate scores are considered, a completely different picture would emerge: all maps could be considered as hierarchical since most of the relations between concepts are held in superordinate/ subordinate relation.

A closer examination of the criteria used to define "hierarchiness" and further analysis of the students' maps (e.g., analysis of the *levels* characteristics in the students' maps) are needed before any final decision is made about the use of hierarchical instructions for constructing concept maps.

Table 11

Means and Standard Deviations of the "Hierarchiness" Scores on Each Group Considering Only the Top Four Students

	Beginner nodes		Cycles		Circuits		Subordinate	
Group	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1 HT/HI	3.13	(1.32)	0	(0)	2.50	(1.29)	.82	(.08)
2 HT/NHI	3.62	(2.28)	0	(0)	1.75	(.50)	.83	(.04)
3 NHT/HI	4.13	(1.18)	0	(0)	2.75	(1.70)	.79	(.13)
4 NHT/NHI	3.25	(1.89)	0	(0)	1.87	(1.55)	.89	(.11)

Conclusion

Criteria used to define hierarchiness prevent us from arriving at a final conclusion about the interaction between hierarchical instructions and the structure of the subject matter domain. We recognize that different conclusions about the structure of the topics and the students' map structures could arise if different experts and "hierarchiness" criteria were used.

It may be that an "averaged" experts' structure should be consider for defining the structure of the domain. This may reduce the problem of variability among experts and provide a better picture of the structure of the content domain (e.g., Acton, Johnson and Goldsmith, 1994).

Which criteria (e.g., subordination, characteristics of levels in the structure, hierarchical cluster) should be used to define hierarchy is a compelling research question worth further exploration.

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.

Anderson, R. C. (1984). Some reflections on the acquisition of knowledge. *Educational Researcher, 13*(10), 5-10.

Anderson, T. H., & Huang, S-C. C. (1989). *On using concept maps to assess the comprehension effects of reading expository text* (Tech. Rep. No. 483). Urban-Champaign: University of Illinois, Center for the Studying of Reading. (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 & Technological Education, 10*(1), 37-52.

Baxter, G. P., Glaser, R., & Raghavan, K. (1994). *Analysis of cognitive demand in selected alternative science assessments* (CSE Tech. Rep. 394). Los Angeles: University of California, Center for the Study of Evaluation, Standards, and Student Testing.

Beyerbach, B. A. (1988). Developing a technical vocabulary on teacher planning: Preservice teachers' concept maps. *Teaching & Teacher Education, 4,* 339-347.

Briscoe, C., & LaMaster, S. U. (1991). Meaningful learning in college biology through concept mapping. *The American Biology Teacher, 53,* 214-219.

Bunge, M. (1969). The metaphysics, epistemology and methodology of levels. In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (pp. 17-28). New York: American Elsevier Publishing Company, Inc.

California Department of Education (1990). *Science framework for California public schools: Kindergarten through grade twelve.* Sacramento, CA: California Department of Education.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.

Dochy, F. J. R. C. (1996). Assessment of domain-specific and domain-transcending prior knowledge: Entry assessment and the use of profile analysis. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning process and prior knowledge* (pp. 93-129). Boston, MA: Kluwer Academic Publishers.

Dress, A., & von Haeseler, A. (1990). Trees and hierarchical structures. In S. Levin, *Lecture Notes in Biomathematics, 84.* Berlin: Springer-Verlag.

Fisher, K. M. (1990). Semantic networking: The new kid on the block. *Journal of Research on Science Teaching, 27,* 1001-1018.

Glaser, R. (in press). Changing the agency for learning: Acquiring expert performance. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts, and sciences, sports, and games.* Mahwah, NJ: Lawrence Erlbaum Associates.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40,* 631-66.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83*(1), 88-96.

Green, M. (1969). Hierarchy: One word, How many concepts? In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (pp. 56-58). New York: American Elsevier Publishing Company, Inc.

Harnisch, D. L., Sato, T., Zheng, P., Yamaji, S. & Connell, M. (in press). Concept mapping approach and its implications in instruction and assessment. *Computers in School.*

Holley C. D., & Danserau, D. F. (1984). The development of spatial learning strategies. In C. D. Holley & D. F. Danserau (Eds.), *Spatial learning strategies. Techniques, applications, and related issues* (pp. 3-19). Orlando, FL: Academic Press.

Hoz, R., Tomer, Y., & Tamir, P. (1990). The relations between disciplinary and pedagogical knowledge and the length of teaching experience of biology and geography teachers. *Journal of Research in Science Teaching, 27,* 973-985.

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

Jones, R. G. (1969). Structural hierarchy: The problem of the one and the many. In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (pp. 56-58). New York: American Elsevier Publishing Company, Inc.

Lay-Dopyera, M., & Beyerbach, B. (1983). *Concept mapping for individual assessment.* Syracuse, NY: Syracuse University, School of Education. (ERIC Document Reproduction Service No. ED 229 399)

Lomask, M., Baron, J. B., Greig, J., & Harrison, C. (1992, March). *ConnMap: Connecticut's use of concept mapping to assess the structure of students' knowledge of science.* Paper presented at the annual meeting of the National Association of Research in Science Teaching, Cambridge, MA.

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*(1), 91-101.

McClure, J. R., & Bell, P. E. (1990). *Effects of an environmental education-related STS approach instruction on cognitive structures of preservice science teachers.* University Park: Pennsylvania State University, College of Education. (ERIC Document Reproduction Service No. ED 341 582)

Mesarovic, M. D., & Macko, D. (1969). Foundations for a scientific theory of hierarchy systems. In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (pp. 29-50). New York: American Elsevier Publishing Company, Inc.

Nakhleh, M. B., & Krajcik, J. S. (1991). *The effect of level of information as presented by different technology on students' understanding of acid, base, and pH concepts.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Lake Geneva, WI. (ERIC Document Reproduction Service No. ED 347 062)

Novak, J. D., & Gowin, D. R. (1984). *Learning how to learn.* New York: Cambridge University Press.

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

Pankratius, W. (1990). Building an organized knowledge base: Concept mapping and achievement in secondary school physics. *Journal of Research in Science Teaching, 27*, 315-333.

Rosen, R. (1969). Comments on the use of the term hierarchy. In L. L. Whyte, A. G. Wilson, & D. Wilson (Eds.), *Hierarchical structures* (pp. 56-58). New York: American Elsevier Publishing Company, Inc.

Ruiz-Primo, M. A., & Shavelson, R. J. (in press). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching.*

Roth, W. M., & Roychoudhury, A. (1993). The concept map as a tool for the collaborative construction of knowledge: A microanalysis of high school physics students. *Journal of Research in Science Teaching, 30*, 503-534.

Shavelson R. J., Baxter, G. P., & Pine, J. (1991). Performance assessments in science. *Applied Measurement in Education, 4*, 347-362.

Shavelson R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology, 63*, 225-234.

Schmid, R. F., & Telaro, G. (1990). Concept mapping as an instructional strategy for high school biology. *Journal of Educational Research, 84*(2), 78-85.

Stice, C. F., & Alvarez, M. C. (1987). Hierarchical concept mapping in the early grades. *Childhood Education, 64*, 86-96.

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*, 1033-1052.

White, R. T., & Gunstone, R. (1992). *Probing understanding.* New York: Falmer Press.

Whyte, L. L., Wilson, A. G., & Wilson, D. (Eds.). (1969). *Hierarchical structures.* New York: American Elsevier Publishing Company.

Wilbraham, A. C., Staley, D. D., Simpson, C. J., & Matta, M. S. (1990). *Chemistry.* Menlo-Park, CA: Addison-Wesley.

Willerman, M. & Mac Harg, R. A. (1991). The concept map as an advance organizer. *Journal of Research in Science Teaching, 28*, 705-711.

APPENDIX A

Compiling the List of Concepts

Teachers were asked to answer these two questions about the unit: (a) "Explain in a few words what you want your students to know when they finish this chapter. In answering this question, think about why this chapter is included in the curriculum and what is the most important thing you want the students to learn about the topic"; and (b) "Based on your answer to question 1, please review the chapter and list all the concepts you think students should know and understand after studying this topic."

Teachers' answers to question 1 involved two aspects of the unit: (a) *conceptual understanding* (e.g., "Students should have a good understanding of the formation of ions, the differences between molecules/compounds . . . molecular/ ionic compounds and acids"); and (b) *application* (e.g., "They should be able to form ionic compounds, binary, ternary, and acids . . . be familiar with the periodic table to identify metals, non-metals. . . ." "Students . . . should be able to write chemical/ molecular formulas; name different substances . . ."). We focused on the conceptual understanding of the unit since concepts maps are about the interrelatedness of concepts. From teachers' responses we concluded that the conceptual understanding about "chemical compounds" (i.e., the other group of substances that are not elements) was the main focus of the unit—how compounds are formed, the types of compounds, and how they can be combined are issues discussed in the chapter.

Answers to question 2 led to a list of 12 concepts considering both teachers' responses (see Table A1 below). We noticed, however, that even though teachers included in their answers to question 1 concepts such as binary ionic compounds or polyatomic ions, they did not include them in the list of concepts in question 2.

Our list included 23 key concepts selected from the chapter (see Table A1). We gave this list to the teachers with the following instructions: "This is a list of concepts that were selected from the chapter 'Chemical Names and Formulas.' Based on what you think are the most important ideas for students to understand about 'Chemical Names and Formulas,' check (√) the concepts that are essential. Please feel free to add any concepts that are missing." Only one of the two teachers returned the list reviewed. Based on the concepts selected by the teacher, we reduced the list to 20 concepts (see Appendix B). This list of 20 concepts was considered to represent the “key/core concepts” of the chapter.

Table A1

List of Concepts Selected From the Revision of the Chapter, The Teachers, and The Expert

Original key/core concept list	Teachers’ list	Expert’s list
1. acids	1. acids	1. acids
2. anions	2. anions	2. anions
3. atoms	3. cations	3. atoms
4. bases	4. compounds	4. bases
5. binary ionic compounds	5. element	5. binary ionic compounds
6. binary molecular compounds	6. ionic charge	6. cations
7. cations	7. ionic compounds	7. compounds
8. compounds	8. molecules	8. electrons
9. electrons	9. molecular compounds	9. elements
10. elements	10. periodic table	10. ions
11. ions	11. chemical formulas	11. ionic compounds
12. ionic compounds	12. molecular formulas	12. metals
13. metals		13. molecules
14. metalloids		14. molecular compounds
15. molecules		15. negative charge
16. molecular compounds		16. neutral charge
17. negative charge		17. non-metals
18. neutral charge		18. representative elements
19. non-metals		19. polyatomic ions
20. polyatomic ions		19. positive charge
21. positive charge		20. ternary ionic compound
22. ternary ionic compound		21. transition elements
23. transition metals		

APPENDIX B

List of Concepts Considered for the Three Conditions

Key/core concept list	List A	List B
1. acids	1. acids	1. acids
2. anions	2. anions	2. anions
3. atoms	3. cations	3. binary ionic compounds
4. bases	4. compounds	4. compounds
5. binary ionic compounds	5. electrons	5. electrons
6. binary molecular compounds	6. ions	6. ions
7. cations	7. metals	7. molecules
8. compounds	8. molecules	8. negative charge
9. electrons	9. molecular compounds	9. non-metals
10. ions	10. polyatomic ions	10. ternary ionic compound
11. ionic compounds		
12. metals		
13. molecules		
14. molecular compounds		
15. negative charge		
16. neutral charge		
17. non-metals		
18. polyatomic ions		
19. positive charge		
20. ternary ionic compound		

APPENDIX C

Sample of Instructions

Instructions for Concept Mapping Technique 1—No Concepts Are Provided to the Students

Name ____ Period ______________________

You recently studied the chapter on Chemical Names and Formulas.

Construct a concept map that reflects what you know about <u>Ions, Molecules, and Compounds</u>.

The concept map should have 10 concepts in it. We are providing you with 3 concepts: ions, molecules, and compounds.

Select another 7 concepts to construct your map. The 7 concepts should be the ones that you think are the most important in explaining ions, molecules, and compounds.

Organize the terms in relation to one another in any way you want. Draw an arrow between the terms you think are related. Label the arrow using phrases or only one or two linking words.

You can construct your map on the blank pages attached. When you finish your map check that: (1) all the arrows have labels; (2) your concept map has 10 concepts, and (3) your map shows what you know about <u>ions, molecules, and compounds</u>.

After checking your map, <u>redraw</u> it so someone else can read it. Staple your <u>final map</u> to this page.

Instructions for Concept Mapping Technique 2—List of 10 Concepts (Sample A) Are Provided to the Students

Name ____ Period ______________________

Examine the concepts listed below. They were selected from the chapter on Chemical Names and Formulas that you recently studied. The terms selected focus on the topic <u>Ions, Molecules, and Compounds</u>.

Construct a concept map using the terms provided below.

Organize the terms in relation to one another in any way you want. Draw an arrow between the terms you think are related. Label the arrow using phrases or only one or two linking words.

You can construct your map on the blank pages attached. When you finish your map check that: (1) all the arrows have labels; (2) your concept map has 10 concepts, and (3) your map shows what you know about <u>ions, molecules, and compounds</u>.

After checking your map, <u>redraw</u> it so someone else can read it. Staple your <u>final map</u> to this page.

You have 30 minutes to construct the map.

LIST OF CONCEPTS
acids
anions
cations
compounds
electrons
ions

metals
molecules
molecular compounds
polyatomic ions

APPENDIX D

Matrix of the Relations Between Pairs of Concepts

	acids	anions	atoms	bases	binary ionic compound	binary molecular compound	cations	compounds	electrons	ions	…	positive charge	ternary ionic compounds
acids	X	1	M* 2	3	M 4	5	M 6	M 7	8	9	…	18	M 19
anions		X	20	21	M 22	23	24	25	M 26	M 27	…	36	37
atoms			X	38	39	40	41	M 42	M 43	M 44	…	53	54
bases				X	55	56	M 57	M 58	59	60	…	69	M 70
binary ionic compound					X	71	M 72	73	74	75	…	84	85
binary molec. compound						X	86	87	88	89	…	98	99
cations							X	100	M 101	M 102	…	111	M 112
compounds								X	113	M 114	…	123	124
electrons									X	M 125	…	134	135
ions										X	…	144	145
⋮											…	⋮	⋮
positive charge												X	190
ternary ionic compounds													X

*M stands for mandatory propositions.

APPENDIX E

Example of the Scoring System for Concept Sample A

LIST A SCORING FORM

Name ______________________________ ID ______________________________

Date ______________________________ Rater ______________________________

Mandatory			**Possible**			**Possible-Less Likely**			**Possible-Unlikely**		
#R	CS	SS	#R	CS	SS	#R	CS	SS	#R	CS	SS
1	4		9	4		8	4		13	4	
6	4		24	4		11	4		29	4	
7	4		25	4		12	4		31	4	
17	4		35	4		30	4		106	4	
26	4		100	4		105	4		139	4	
27	4		110	4		128	4		156	4	
101	4		113	4		133	4				
102	4		116	4		138	4				
104	4		122	4		155	4				
114	4		127	4		160	4				
117	4		129	4		167	4				
118	4		137	4		173	4				
125	4										
143	4										
163	4										

Concepts on List:
acids ______
anions ______
cations ______
compounds ______
electrons ______
ions ______
metals ______
molecules ______
mol comp. ______
polya. ions ______

Comments: ______________

Other Concepts: ______________

\# of Concepts: ______/______

\# Valid Propos: ______________

\# of Propositions: ______________

Total **M** ☐ + Total **P** ☐ +Total **LL** ☐ +Total **U** ☐ Grand Total ☐

APPENDIX F

Atom Structure Concept Lists: Key/Core Concept List, Researchers' List, Teachers' List, and Expert's List

Key/core concept list	Researchers' list	Teachers' list	Physicist's list
1. atom	1. atom	1. atom	1. atom
2. atomic mass	2. atomic mass	2. atomic mass	2. atomic mass
3. atomic number	3. atomic number	3. atomic number	3. atomic number
4. atomic orbitals	4. electron	4. *d* orbital	4. binding energy
5. electron	5. isotope	5. Dalton's atomic theory	5. electromagnetic force
6. elements	6. mass number	6. electron	6. electron
7. energy levels	7. negative charge	7. element	7. filled orbitals
8. isotope	8. neutral charge	8. energy levels	8. ions
9. mass number	9. neutron	9. isotope	9. isotope
10. negative charge	10. nucleus	10. mass number	10. mass number
11. neutral charge	11. orbitals	11. negative charge	11. neutron
12. neutron	12. positive charge	12. neutral charge	12. nucleus
13. nucleus	13. proton	13. neutron	13. electron cloud (orbitals)
14. *p* orbitals	14. subatomic particles	14. nucleus	14. Pauli exclusion principle
15. positive charge		15. orbitals	15. periodic table
16. proton		16. *p* orbitals	16. proton
17. *s* orbitals		17. periodic table	17. photoelectric effect
		18. positive charge	18. quarks
		19. proton	19. shape of orbitals
		20. *s* orbitals	20. strong force
			21. subatomic particles
			22. unfilled orbitals
			23. weak force

APPENDIX G

Sample of Hierarchical and Nonhierarchical Instructions

Concept Mapping Technique 1 Instructions—Hierarchical Structure is Imposed

Name____ Period _______

Examine the concepts listed below. They were selected from the chapter on Atomic Structure that you recently studied. Construct a hierarchical concept map using the terms provided below. Organize more general terms above the more specific ones. Draw a line between the terms you think are related. Label the line using phrases or only one or two words.

You can construct your map on the blank pages attached. When you finish your map check that: (1) you have all the concepts on the list in your map; (2) all the lines have labels; (3) your map is explaining atomic structure. After checking your map, <u>redraw</u> it so someone else can read it.

Staple your <u>final map</u> to this page.

LIST OF CONCEPTS
atoms
atomic mass
atomic number
atomic orbitals
electrons
elements
energy levels
isotopes
mass number
negative charge
neutral charge
neutrons
nucleus
p orbitals
positive charge
protons
s orbitals

Concept Mapping Technique 2 Instructions—No Specific Structure is Imposed

Name____ Period _______

Examine the concepts listed below. They were selected from the chapter on Atomic Structure that you recently studied. Construct a concept map using the terms provided below. Organize the terms in relation to one another in any way you want. Draw a line between the terms you think are related. Label the line using phrases or only one or two words.

You can construct your map on the blank pages attached. When you finish your map check that: (1) you have all the concepts on the list in your map; (2) all the lines have labels; (3) your map is explaining atomic structure. After checking your map <u>redraw</u> it so someone else can read it.

Staple your <u>final map</u> to this page.

LIST OF CONCEPTS
atoms
atomic mass
atomic number
atomic orbitals
electrons
elements
energy levels
isotopes
mass number
negative charge
neutral charge
neutrons
nucleus
p orbitals
positive charge
protons
s orbitals

APPENDIX H

ATOM STRUCTURE SCORING FORM

Name ________________________________ ID ________________________________

Date ________________________________ Rater ________________________________

Mandatory			Possible			No-Relation			Forbidden		
# R	VS	IS	# R	VS	IS	# R	VS	IS	# R	VS	IS
2			1			17			12		
3			4			21			14		
5			6			25			64		
7			8			26			68		
11			9			27			77		
22			10			28			79		
33			13			29			111		
35			15			31			113		
40			16			36			116		
42			18			37			117		
45			19			38			119		
49			20			39			120		
57			23			41			122		
60			24			43			124		
66			30			44			129		
72			32			51			131		
92			34			52			135		
93			46			53					
99			47			54					
101			48			56					
107			50			61					
109			55			62					
110			58			63					
125			59			82					
134			65			83					
			67			84					
			69			85					
			70			86					
			71			87					
			73			89					
			74			90					
			75			94					
			76			96					
			78			97					
			80			98					
			81			100					
			88			102					
			91			104					
			95			105					
			103			106					
			123			108					
			126			112					
			127			114					
			130			115					
			133			118					
						121					
						128					
						132					
						136					

Comments:

____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____

# of Concepts: ________

atoms ________
atomic orbitals ________
atomic mass ________
atomic number ________
electrons ________
elements ________
energy level ________
isotopes ________
mass number ________
neutrons ________
neutral charge ________
nucleus ________
negative charge ________
p orbitals ________
positive charge ________
protons ________
s orbitals ________

Valid Prop ________

Total # of Prop ________

Incusive Rel. ________

of Cycles ________

# of Circuits ________

# of Beg Nodes ________

Total M ☐ Total P ☐ Total NR ☐ Total F ☐
Grand Total ☐