ESTIMATING THE RELIABILITY AND VALIDITY OF CONCEPT MAPPING AS A TOOL TO ASSESS PRIOR KNOWLEDGE

By

Susan Lee Coleman

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Instructional Technology and Design

APPROVED:

Thomas M. Sherman, Chair

John K. Burton

Elizabeth Koball

Marcia N. Murawski

D. Mike Moore

November, 1994

Blacksburg, Virginia

ESTIMATING THE RELIABILITY AND VALIDITY OF CONCEPT MAPPING AS A TOOL TO ASSESS PRIOR KNOWLEDGE

by

Susan Lee Coleman

Thomas Sherman, Chair
Curriculum and Instruction

(ABSTRACT)

An initial activity in the instructional design process is an assessment of learner characteristics such as prior knowledge. Although this step is important for the instructional design process, Reiss and Morley (1993) found that only about half of the practicing instructional technologists they surveyed always performed some kind of learner analysis. The reasons for not conducting a learner analysis included lack of time and lack of knowledge about how to do such an analysis.

Concept mapping has been proposed as a convenient tool for assessing learners' prior knowledge (Jonassen, Beissner, & Yacci, 1993; Novak & Gowin, 1993). It has been said that a concept mapping procedure can be quickly developed and then conveniently administered to learners (Jonassen et al., 1993; Novak & Gowin, 1993). However, there is no evidence supporting the reliability or validity of concept mapping as a way to measure learners' prior knowledge. The purpose of this research was to investigate the reliability and validity of concept mapping as a tool to assess learners' prior knowledge.

Students in a graduate-level statistics class and an undergraduate level engineering class were administered concept mapping and multiple-choice tests before and after instruction. The Interrator reliability, internal consistency, content validity, face validity, concurrent validity, and construct validity of the concept mapping procedure were then examined.

Overall, the data do not support the reliability and validity of the concept mapping procedure used in this study. However, the reliability and validity levels achieved by the concept mapping test were comparable to the multiple-choice test developed by the classroom instructor. It may be that concept mapping is no less reliable and valid than other forms of classroom assessments. Nonetheless, the concept mapping procedure used in this study can not be recommend as a practical measure of students' prior knowledge. In addition to the poor reliability and validity, the concept mapping procedure is difficult to learn and the scoring procedure is difficult to use. In the end, the user will have to weigh the results of this study and judge the adequacy of concept mapping for its intended purpose.

Acknowledgments

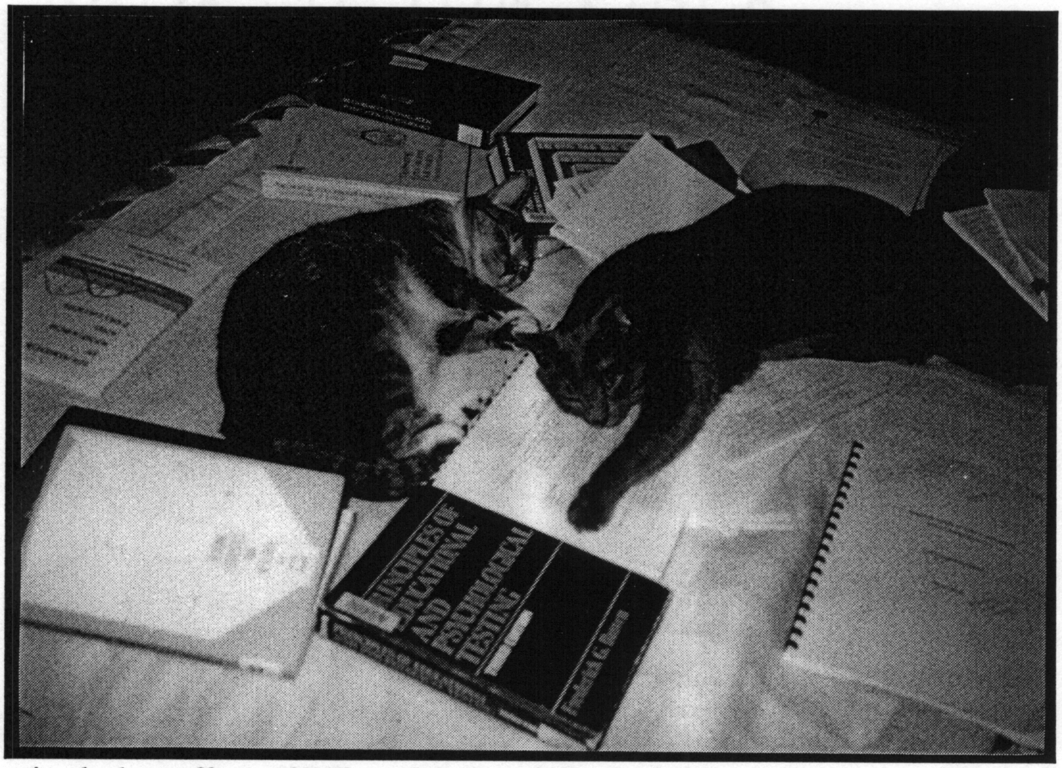
A Ph.D. cannot be accomplished without a strong support system backing up the Ph.D. candidate. I had one of the best. My advisor, Tom Sherman, encouraged me to begin a Ph.D. program and never lost confidence in me. I will always be grateful for his unwavering support. My committee could not have been more supportive. Mike Moore and John Burton treated me as much like a colleague as a graduate student. Betty Koball was more than my "statistics guru," she was my friend and helped me through some difficult times. Marci Murawski has always been a role model for me. She demonstrates what a woman with style and acumen can accomplish as an instructional technologist. There is much more I want to learn from her.

I want to thank Eileen Van Aken for helping me with my data collection and ranking concept maps. I also want to thank my "editor," Susan Bixler, who read every page of my dissertation to make sure my apostrophes and commas were all in the right places.

To all of my co-workers, I thank you for your patience and understanding when I had a paper due or had to miss work to attend a class. I particularly want to thank Management Systems Laboratories for giving me the opportunity to practice instructional design for over three years while I worked on my dissertation.

My family was not always sure what I was doing in school for so long, but they always supported my education. I want to thank my mom for having faith in me, and my brother Dean for living in Utah and providing me a great place to ski during my breaks. I dedicate my dissertation to the memory of my dad. I know how proud he would have been of me and I know he will be with me in spirit at my graduation.

Billy and Jesse always have been and always will be a constant source of comfort and/or stress. Our boyz were always there to lend a helping paw, hoping that I would soon finish my studies so that I once again could provide them with my full, undivided, and constant attention.



Despite the best efforts of Billy and Jesse to help me finish my dissertation, I finished anyway. Thanks boyz.

And finally, I would not and could not possibly be where I am today without the love and support of my husband, Garry. My support system wouldn't have been enough without the steadfast commitment and devotion Garry gave me. You are forever in my heart.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Ausubel (1968) states, "The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly" (p. vi). Many instructional design models reflect this theory and guide instructional designers to assess learners before designing instruction. However, current procedures available for determining what learners already know are not always practical when solving instructional design problems in the real-world. This study will examine concept mapping as a procedure to determine learners' prior knowledge. In particular, the validity and reliability of concept mapping as a measure of learners' prior knowledge will be investigated.

Assessing prior knowledge for instructional design

A common beginning point in most instructional design models is a target-audience analysis (Richey, 1986). For example, in the Dick and Carey (1985) model, instructional designers determine learners' entry-level characteristics during the second phase of the model. The results of a target-audience analysis provide information regarding the entry level knowledge, skills, and attitudes of the target audience and are often applied to instructional decision making (Seels & Glasgow, 1990).

There are many learner characteristics that may be appropriate to assess before instruction. Prior knowledge, motivation to learn, anxiety about learning, verbal abilities, and field dependence may be relevant to instruction (Seels & Glasgow, 1990). The selection of which characteristics to assess will depend on the learning situation (Dick & Carey, 1985) and how the information will be used. Prior knowledge, or learners' current understandings of concepts related to a domain of knowledge, is generally considered essential (e.g., Gagné, Briggs, & Wager, 1988). An analysis of learners' prior knowledge is useful to an instructional designer for several reasons.

Establishing learners' prior knowledge allows designers to identify knowledge gaps to be filled, misconceptions to be corrected, and concepts to be elaborated upon. Knowing these things, instructional designers then have some idea of where instruction should begin (e.g., start with the basics or begin with more complex ideas), what topics to include during instruction, how to present these topics (i.e., the selection of appropriate instructional activities) (Sax, 1989), and any learner misconceptions that must be unlearned before learning may progress. For example, an assessment of learners' prior knowledge may indicate learners generally have a strong understanding of basic concepts in a domain, but have limited understanding of the relationships among those concepts. The instructional designer may determine that these learners will benefit by analyzing a series of case studies and consequently include case studies in the instructional strategy for the course.

Establishing prior knowledge provides information that can be used as part of summative and formative evaluations to determine the success of the learning program (Camp et al, 1986). Pre- and post-evaluation of learners' knowledge indicates growth during a course and may be used as a summative evaluation to assess the effectiveness of instruction and the level of learners' achievement. Using prior knowledge as a standard to compare changes during instruction allows instructors or instructional designers to perform formative evaluations that may result in mid-course changes in instruction if learners are not progressing as expected.

Establishing prior knowledge can provide information to instructional designers to select learners for instructional programs. Frequently, prerequisite knowledge is necessary for successful learning in some domains of knowledge. For example, understanding simple geometry may be necessary for learning to survey land. Assessing learners' prior knowledge can indicate whether learners have the geometry knowledge base to learn the subject matter addressed in a course. The instructional designer may want to discourage or disqualify learners who do not possess the prerequisite knowledge.

While understanding learners' prior knowledge is accepted as important, a recent study by Reiss and Morley (1993) suggests what is best is not always what is done. Reiss and Morley surveyed members of the National Society of Performance and Instruction to determine what steps in the instructional systems design process are typically skipped by practicing instructional designers and to determine why these steps are skipped. They found that only 51% of the respondents always performed some kind of in-depth analysis of the target audience before developing instruction. Furthermore, Reiss and Morley reported that of the reasons for not conducting an analysis of the target audience, 44% of the respondents reported a lack of time and 12% reported a lack of knowledge about how to do such an analysis.

Perhaps measuring instruments or procedures that are quickly developed and used, and provide easily-analyzed data could make formal target audience assessment, and particularly, prior knowledge assessment more accessible to practicing instructional designers. There are a number of informal and formal methods for assessing prior knowledge. Informal methods may include observation, informal interviews, and interaction with members of the intended audience. Formal methods may include essay and multiple-choice tests (Mehrens & Lehmann, 1991). A popular technique for formally measuring prior knowledge is the multiple-choice test (Mehrens & Lehmann, 1991; Popham, 1981). Although well-constructed multiple-choice tests can provide a great deal of information about what learners know, constructing, administering, and interpreting the results of a valid and reliable multiple-choice test is difficult and time consuming, especially for novices (Mehrens & Lehmann, 1991; Sax, 1989).

Concept Mapping procedures, producing representations of learners' prior knowledge, are typically uncomplicated to develop and implement, and provide data that are readily analyzed (Novak & Gowin, 1983). A Concept Mapping procedure may also prove to be a reliable and valid way to assess prior knowledge. There are many different Concept Mapping procedures (e.g., the word association method, interview method, and graph construction methods). One specific graph construction method is called concept mapping. This is not to be confused with Concept Mapping which is the general term for all procedures to graphically portray student knowledge (in this paper, Concept Mapping will be capitalized when referring to the general term). The concept mapping procedure, as developed by Novak and Gowin (1993), will be the focus of this study.

The purpose of this study is to investigate concept mapping to estimate its validity and reliability as part of a target-audience analysis to determine students' prior knowledge. As the Reiss and Morley study (1993) suggests, practicing instructional designers often appear to avoid target audience analysis because they have neither the time nor the expertise to carry out the assessment. If reliable and valid, concept mapping could be an effective and convenient assessment tool for instructional design professionals.

In the next section, several Concept Mapping procedures are discussed and the rationale for selecting concept mapping as the focus of this study is presented. Then, a discussion of ways concept mapping and other Concept Mapping procedures have been used in research and practice is presented to show the need for reliability and validity studies on Concept Mapping procedures.

Techniques for representing cognitive structure

A basic tenet of cognitive psychology is that learners' knowledge is organized into an interrelated network of concepts that differs among individuals. This organization of knowledge has been called cognitive structure (Ausubel, 1968; Jonassen et al., 1993; Shavelson, 1972), schemata (Anderson, 1977), and frames (Minsky, 1975). For the purpose of this research, the term cognitive structure will be used to describe learners' structure of knowledge. Ausubel (1963) defines cognitive structure as "an individual's organization, stability, and clarity of knowledge in a particular subject matter field at any given time" (p. 26). Shavelson (1972) defines cognitive structure as "a hypothetical construct referring to the organization (relationships) of concepts in memory" (p. 226-227). Jonassen et al. (1993) refer to cognitive structure as structural knowledge or the "pattern of relationships among concepts in memory" (p. 4)

Concept Mapping techniques attempt to represent learners' cognitive structure related to a domain of knowledge. Each Concept Mapping technique is unique in the way learners state the concepts they see as related to the target concept, and in the way the relationships among the concepts are portrayed. Several Concept Mapping procedures have been developed as techniques for learners to represent their prior knowledge.

Concept Mapping procedures share the problem of trying to portray cognitive structure; a structure we are unable to see. However, none of the developers of these Concept Mapping procedures claim to produce a replication of cognitive structure, only a representation. The final representation of cognitive structure reflects the Concept Mapping technique used to create that representation (Donald, 1987). When Concept Mapping techniques are used to represent cognitive structure, the user must remember they are viewing a representation based on the selected Concept Mapping technique and not a replication of learners' cognitive structures.

Another problem with trying to reveal learners' cognitive structure is that Concept Mapping procedures impose structure to the content being mapped. Some domains are less structured than others, and when structure is imposed on ill-structured content areas by a Concept Mapping procedure, the learner may find it difficult to demonstrate understanding through concept mapping. For example, counseling is a somewhat ill-structured content area (e.g., causal relationships are not always clear and some terminology is indefinite). Some basic theories in counseling may be mappable, but using Concept Mapping to demonstrate an understanding of counseling may be difficult.

A last problem with Concept Mapping is there are many learner characteristics that may affect learners' abilities to map a topic. For example, McClure (1989) found concrete operational and formal operational subjects mapped geometric concepts differently. This will be discussed in more detail in a later section. Other possible intervening variables might be field dependence, creativity, verbal skills, writing skills, and the learners' ability to use mnemonics. Few studies appear in the literature that look at how different learner characteristics affect learners' abilities to complete a Concept Mapping technique.

Despite the difficulty of trying to represent cognitive structure, Concept Mapping procedures continue to demonstrate merit as a procedure to measure learners' cognitive structure. Examples of research and practical applications of Concept Mapping will be discussed in a later section. First, several procedures that show promise as effective ways to portray learners' cognitive structure are presented.

Word association technique

The word association technique requires learners to list words they consider related to the main concept of interest. A key word is provided representing the main idea of the concept being mapped. This word is written in a column on the left side of a page. Space is provided to the right of each word for learners to write words related to the main idea. Repeating the main idea down the left side of the page orients the learner to the main idea each time another word is written. The word association method can be controlled or the student can be allowed to free associate. A list of stimulus words are provided under the controlled conditions and the learner simply chooses words from that list in order of importance (see Figure 1). No such words are provided during the free-association method and learners simply brainstorm words (see Figure 2) (Jonassen et al., 1993). The number, kind, order, and overlap of responses are considered in the analysis of the raw data gathered from word-association tasks (Shavelson, 1974). The raw data are graphed using multidimensional scaling or cluster analysis.

Protocol analysis method

Representations of what learners know may also be produced from an interview, or protocol analysis. In this case, students do not construct their own representations, but the interviewer infers learners' understanding from the interview data. Raw data gathered through the interview are analyzed by the interviewer for stated relationships, strategies, assumptions, and logical errors (Donald, 1987). Interviewers then construct maps to resemble one of the graph construction methods described below.

Directions: The key word "politician" is printed down the left side of this page. A list of words related to the concept of a politician appears below these directions. Think about the concept of a politician, then go to the word list and choose the word you think is most related to the key word. Write that down on the first blank. Read the key word again and go back to the word list to choose the next word most closely related to your idea of a politician. Write that word in the next blank space. Repeat this process until all the words have been selected from the word list.

representative	speeches	lawyer	campaign	house	votes
debate	senate	election	laws	promises	amendments

politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________
politician ______________________________

Figure 1. Example - Controlled word association procedure

Directions: The key word "politician" is printed down the left side of this page. Each time the word appears, read it, then think about what it means. Write the first word that comes to mind on the blank next to the key word. Continue doing so until you can't think of any more words.

politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________
politician ______________________

Figure 2. Example - Free word association procedure

Graph construction methods

Graph construction methods result in a representation of cognitive structure in which relationships between concepts are specified by connecting lines. As with the word association method, there are variations including tree construction, pattern note, and concept mapping procedures.

During the tree construction procedure, subjects are given a list of words from which they first choose the two words most closely related. These two words are written on a blank sheet of paper and connected with a line labeled with the number one. Students then look back at the list of words, find the word most closely related to either of the first two words, and write that word next to the related word on the page. A line is drawn to connect these two words and labeled with the number two. This continues until all the words in the list have been connected. The result is an arrangement of words connected with numbered lines representing the order in which the words were connected (see Figure 3). The tree-construction procedure works particularly well with well-structured disciplines such as chemistry (Donald, 1987).

representative politician lawyer campaign house votes
debate senate election laws promises amendments

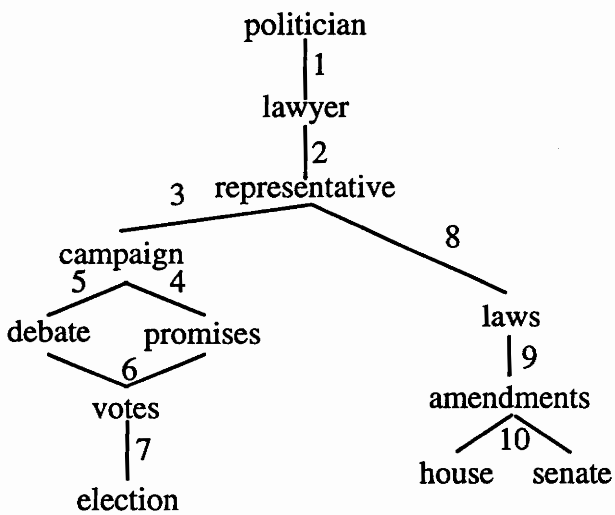


Figure 3. Example - Tree construction

The pattern note procedure gives few cues to students. During this procedure, learners are given the topic to be mapped and are instructed to come up with one word representing the main idea of the topic. They write this word on a sheet of paper and draw a box around it. Learners free associate about the topic and describe their initial ideas on lines connected to the box. Secondary or tertiary thoughts are written on lines connected to the initial ideas. This continues until students feel they have adequately represented the main idea. Relationships between subordinate ideas can be represented with additional connecting lines (see Figure 4).

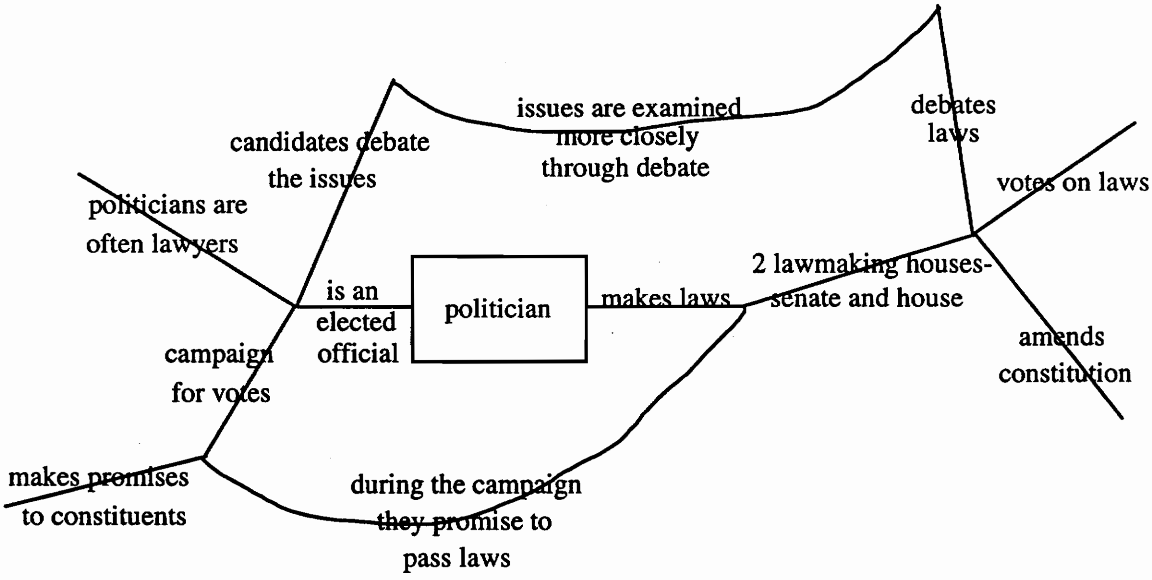


Figure 4. Example - Pattern note

Concept mapping was developed by Novak and Gowin (1993) to graphically portray what learners know about a given domain. A concept is defined as "a regularity in events or objects designated by some label" (Novak & Gowin, 1993, p. 2). Concept maps are made up of propositions which are composed of two or more concepts linked by words describing the relationship between the two concepts (Novak & Gowin, 1993). For example in the proposition "football is a sport", "football" and "sport" are two concepts; the relationship between these concepts is described by the phrase "is a".

Concept maps are hierarchically organized with the most inclusive concept placed at the top of the map and subordinate concepts appearing below (see Figure 5). Learners are given a topic and a list of words representing concepts related to that topic. (Another option is to ask learners to brainstorm a list of words they believe to be related to the topic being mapped.) Then they list the words in order of importance. From this list they choose the most inclusive concept, write it down on a piece of paper, and circle it. They return to the list, write the second word under the first word that was circled, draw a line between them, and describe the relationship between these words on the line. This is repeated until all the words from their lists are gone. Learners then study the resulting concept map and look for additional relationships among concepts that have not yet been shown on their maps (Jonassen et al., 1993). The concept mapping procedure allows students to graphically portray relationships they understand to exist among many concepts within a domain of knowledge. Concept maps can be scored using the guidelines provided by Novak and Gowin (1993) (see Appendix 1), or as Novak and Gowin also suggest, they can simply be reviewed for the qualitative information they contain.

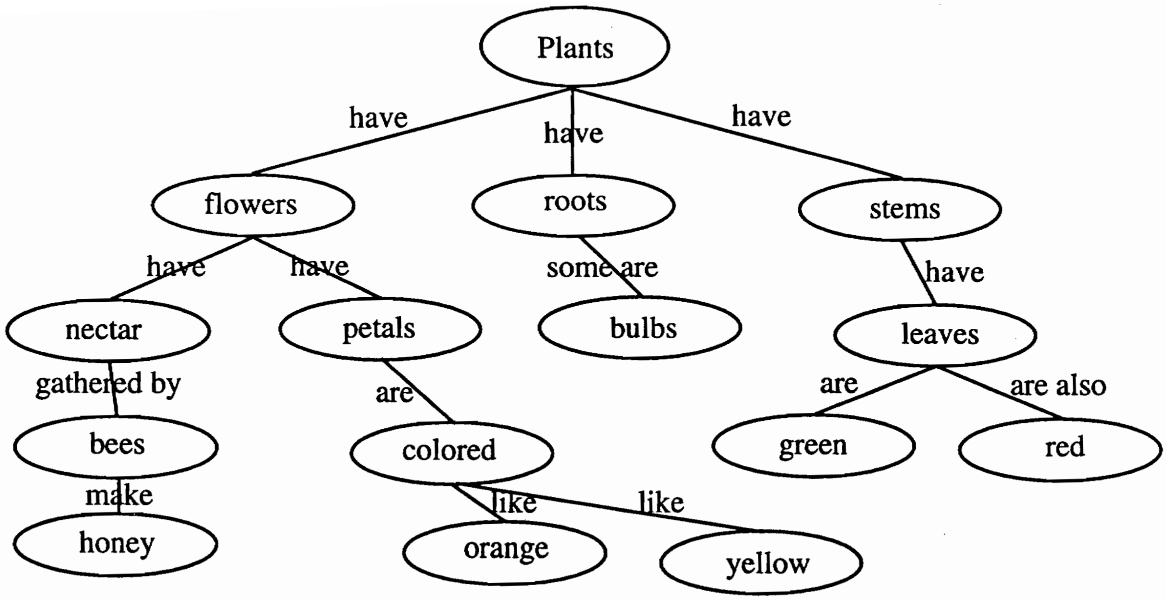


Figure 5. Example - Concept map (from Novak & Gowin, 1993, p. 176)

<u>Applications of Concept Mapping procedures in research and practice</u>

The purpose of this section is to describe ways Concept Mapping has been used in research and practice to help improve instruction. The absence of reliability and validity estimates of the Concept Mapping procedure used in each study will also be demonstrated.

McDonald (1989) demonstrated that Concept Mapping may be useful for instructional designers to predict the preparedness of students to understand complex ideas. McDonald used a variation of the word association technique as a research tool to investigate geometric concepts of students in different developmental stages. McDonald found that concrete operational students (able to understand logical operations with concrete things only) represented geometric concepts only superficially, while formal operational students (able to think logically with abstrations) constructed more stable and accurate representations of their knowledge structures. Data for this study were collected with a variation of the word association technique in which the students were asked to rank the degree of the relationship between pairs of words from one to ten. The data were then graphed using multidimensional scaling.

After analyzing the graphs representing students' knowledge, McDonald concluded that concrete operational students may not be ready to understand abstract concepts, and the instructional strategies chosen to teach abstract concepts may need modification to accommodate these students. Thus, McDonald found Concept Mapping could provide instructional designers information about students that will help them select more effective and efficient instructional methods. The reliability and validity of this Concept Mapping procedure were not reported in McDonald's study.

Research by Lewis, (1991), Gussarsky and Gorodetsky (1988), Shavelson (1972), and Beyerbach and Smith (1992) illustrate how Concept Maps may be used for student assessment before, during, or after instruction.

Lewis (1991) demonstrated Concept Mapping reveals a range of learners' abilities to process information. Lewis used protocol analysis to gather data to construct representations of eighth grade students' cognitive structures during an elementary thermodynamics class. An analysis of these representations led to the identification of three classifications of students determined by the degree to which they accurately restructured their knowledge as new information was received. Almost one-fourth of the students were described as "converging." These were students who continually modified their knowledge structures based on new information. They understood underlying thermodynamic processes, enabling them to build accurate connections between concepts. More than half the students were in the second category called "progressing." Although these students modified their knowledge structures based on new information, they could not integrate concepts as well as "converging" students. The remaining one-fourth were considered "oscillating" because they changed their concepts sporadically and not necessarily based on new information. These students perceived knowledge at a surface level, not seeing underlying ideas which gave the information deeper meaning. By the end of the course, the oscillating students had not yet built accurate relationships among concepts. Lewis concluded the knowledge structures of the students in these three groups were different, indicating different learning needs. For example, Lewis found oscillating students benefited from modeling problem solving processes. Lewis does not report the reliability and validity estimates for the Concept Mapping procedure used in the study.

Gussarsky and Gorodetsky (1988) demonstrated Concept Mapping procedures can be used to track changes in students' cognitive structure during instruction. Gussarsky and Gorodetsky used a word association method to study how the formation of chemical equilibrium concepts in college students is affected as a chemistry class progressed through a semester. From the data, the researchers were able to identify how students' misconceptions lead to other misconceptions (e.g., misunderstanding the Le Chatelier principle is reflected in the lack of connections between this concept to others, further affecting the understanding of these follow-on concepts). Gussarsky and Gorodetsky also showed changes in the results of the word association exercise may indicate how specific units of information entering learners' knowledge structures affect understanding of existing relationships. Understanding how new information modifies prior knowledge may enable instructional designers to make better decisions regarding the order and presentation methods of instructional material. The reliability and validity of the Concept Mapping procedure were not reported by the authors.

Shavelson (1972) demonstrated Concept Mapping can be used to track changes in cognitive structure over time. Shavelson used the word association technique to investigate the understanding of physics concepts of students in a high school physics course. Data were collected before, during, and after instruction for an experimental group receiving instruction. The same data were collected for a control group not receiving instruction. From the data, Shavelson concluded the cognitive structures of students in the experimental group changed as a result of instruction. Because changes in the cognitive structures of students can be detected over time, and because these changes can be attributed to instruction, the effectiveness of an instructional program may be evaluated by monitoring students' cognitive changes with a Concept Mapping technique. Student achievement may also be evaluated using the same technique. Shavelson did not report the reliability or validity of the Concept Mapping procedure.

Beyerbach and Smith (1992) demonstrated Concept Mapping can be used as an evaluation component for a teacher education program. Beyerbach and Smith compared several different methods for studying preservice teachers' thinking to investigate how their thinking evolved throughout their participation in a teacher education program. One of the approaches the researchers used was concept mapping. Preservice teachers periodically created concept maps for two years. The first year they used paper and pencil to construct concept maps, the second year they constructed maps using a computer program called Learning Tool. The preservice teachers constructed and continually updated concept maps for the topic, effective teaching. Beyerbach and Smith found students' technical vocabulary increased as they progressed through the program. Also, students' maps became more differentiated, more hierarchically organized, and more similar to each others' and to the instructors' maps. The program administrators used the data gathered from this study to make program implementation decisions, such as introducing less technical vocabulary in more depth and including more classroom management topics in the program. The researchers concluded concept mapping provided one perspective of preservice teachers' knowledge, and, in conjunction with the other approaches (stimulated recall, discourse analysis, and reflective journals), the researchers were able to piece together an understanding of preservice teachers' knowledge of effective teaching. This allowed the researchers to make informed decisions regarding the content of the program. Reliability and validity estimates were absent from this study.

McClure and Bell (1990) used concept maps as a measure of college students' cognitive structure before and after completing an environmental education course to determine if the course fulfilled course objectives. The course was specifically designed to help students learn additional science concepts, and build an understanding of how these concepts relate to other non-science areas such as social issues. An analysis of the concept maps before and after instruction revealed that while existing structures were enlarged with new information, there was no increase in propositions connecting concepts in different domains as was expected from the design of the course. No reliability or validity estimates were reported in this study.

Okebukola and Jegede (1988) demonstrated how concept mapping can be used as a tool for teaching. They found when concept mapping was used as a teaching tool in an undergraduate science class, students with a cognitive orientation toward learning principles performed better on an achievement test than students whose learning strategy was characterized by simple recall of information, critical questioning of information, or application of information in problem solving. Also, students in a cooperative learning setting in which concept mapping was used as a teaching tool performed better on an achievement test than students in an individualistic setting. Reliability and validity issues were not discussed in this study.

These studies illustrate how several different Concept Mapping procedures have been used to measure cognitive structure in research and practice. A common problem with the body of literature from which these studies come, is the absence of reliability and validity estimates when Concept Mapping is used as a measuring device. Thorough reliability and validity studies must be pursued before these procedures can be used confidently by instructional designers. This study will investigate the reliability and validity of one Concept Mapping procedure as a way to represent cognitive structure.

An alternative for assessing prior knowledge

Of the many different procedures for representing cognitive structure, concept mapping has been suggested as a way to portray learners' prior knowledge (Clark, 1990; Johnson & Thomas, 1992; Jonassen et al., 1993; Novak & Gowin, 1993). This procedure will be the focus of this reliability and validity study. Concept mapping appears to be particularly useful for five reasons.

First, developing and implementing the procedure is straightforward (Novak & Gowin, 1993). The test developer does not need an extensive background in test construction to use concept mapping. The test developer chooses the concept to be mapped, then lists a number of concepts related to the main concept. Another option is to let learners come up with their own list of concepts related to the main concept, which may provide even richer data because learners are determining the concepts to map (investigations regarding this kind of comparison of concept mapping methods do not appear to exist in the literature).

Second, concept mapping gives learners the freedom to create a representation that fits their understanding of the relationships among concepts. In contrast, the word association procedures ask learners to brainstorm a list of words. Representations from word association procedures are generated from the list through multidimensional scaling or cluster analysis; both analyses are more difficult and time consuming to carry out than scoring a concept map. Learners do not create their own "picture" of their prior knowledge as they do through concept mapping. Furthermore, the graphs produced through multidimensional scaling or cluster analysis do not indicate the relationships learners understand to exist between concepts. During concept mapping, however, learners explain their understanding of relationships between concepts directly on their concept maps by connecting concepts with lines and explaining the relationship between concepts directly on these lines.

Third, although learners have considerable freedom to place concepts where they choose, they are somewhat confined by the concept mapping instructions to produce a hierarchical structure. Furthermore, this kind of structure is rewarded by the scoring algorithm, where each level of a hierarchical structure is awarded five points. This limitation makes scoring concept maps convenient, unlike pattern notes. The pattern note technique permits learners to create their own representation of their prior knowledge without constraints, but the resulting maps are so free-formed that reliable scoring becomes very difficult. However, the concept mapping procedure provides good balance between permitting learners the freedom to choose the contents of their maps, and giving them some boundaries in which to construct their maps.

Fourth, concept maps are constructed by learners, and not by a second party as with representations constructed by interviewers using interview data. When someone else creates a representation of learners' knowledge, there is always a chance the second-party has misrepresented learners' knowledge.

And fifth, the analysis of data gathered from concept mapping is uncomplicated. Novak and Gowin (1993) have developed a scoring procedure for the maps (see Appendix 1), enabling learners' to be ranked by their map scores. Furthermore, simply examining the relationships learners demonstrate among concepts may reveal a great deal about learners' prior knowledge. Jonassen (1987) suggested complicated statistical analyses like multidimensional scaling and cluster analysis (required for the word association methods) discourage many practicing educators from using some Concept Mapping techniques. A procedure like concept mapping may provide data requiring less complicated analysis.

Conceptual basis for concept mapping

David Ausubel's cognitive theory of learning (Ausubel, 1968) provides the basis for Novak and Gowin's (1993) concept mapping procedure. The primary concept in Ausubel's theory is meaningful learning results when new knowledge is subsumed into existing relevant concepts and propositions are organized hierarchically in a cognitive structure. More inclusive concepts appear at the top of this hierarchy and less inclusive and more specific concepts and propositions appear toward the bottom. Novak and Gowin assert that concept maps, being a representation of cognitive structure, should reflect a hierarchical organization of knowledge. Therefore, the concept mapping procedure results in a hierarchical structure, representing how learners organize knowledge.

The concept map scoring algorithm developed by Novak and Gowin (1993) is also based on Ausubel's cognitive theory of learning. First, points are awarded for concept maps showing hierarchies ordered with the most inclusive concepts at the top and the less inclusive concepts below. The hierarchical organization must be accurate for points to be awarded. For example, it would not make sense to have "giraffe" as an inclusive concept, then have the concept "mammals" as a less inclusive concept. Five points are awarded for each relevant level of the hierarchy.

Second, points are awarded for accurate relationships between the hierarchical concepts. Hierarchies change and develop as new information is added to existing cognitive structures in a process called progressive differentiation (Ausubel, 1968). This is a continuous process whereby we add meaning to existing concepts as new relationships are learned (Ausubel, 1968). Existing concepts are modified to reflect the addition or subsumption of new information. Therefore, concepts are never really complete, but are constantly challenged or confirmed based on new information. Concepts are re-examined to more clearly differentiate them from other similar, but different concepts. Also, detail may be added to concepts, further differentiating them from other concepts. The exclusiveness and accuracy of concepts in the concept map hierarchy are suggestive of this process occurring (Novak & Gowin, 1993). The scoring algorithm awards one point for each correct, meaningful relationships demonstrated between two hierarchical concepts.

Third, points are awarded for meaningful cross links between segments of the concept map. Integrative reconciliation is a process during which existing, related concepts are examined for similarities and differences, resulting in the discovery of new relationships among concepts (Ausubel, 1968; Novak & Ridley, 1988). The complexity of the hierarchical structure in a concept map is further developed by recognizing relationships between existing sections of the structure. During integrative reconciliation, students discover new relationships among existing concepts (Ausubel, 1968). Concept maps showing cross links among concepts suggest this process has taken place. Ten points are awarded for each cross link that is both valid and significant. The final scoring algorithm developed by Novak and Gowin (1993) is summarized in Figure 6.

Purpose of this study

Concept mapping is a promising procedure for assessing prior knowledge, but has not been established as reliable and valid. The purpose of this study is to investigate the reliability and validity of the concept mapping procedure to assess prior knowledge.

Research questions

How reliable and valid is concept mapping as a measure of what learners know prior to instruction?

- How much variance in learners' concept map scores are attributable to differences in the scorers?
- To what extent do the items (concepts) on a concept mapping test measure the same characteristic? (This question assumes learners are given a list of concepts to map.)
- How well do the items on a concept mapping test represent the universe of items relevant to the target topic?
- To what extent do learners and course administrators perceive the concept mapping test measures what it is supposed to measure?
- Will scores on concept mapping correlate sufficiently with scores on a multiple-choice test to indicate these two tests measure the same content?
- Will learners' concept maps change after instruction in the target concepts? And if the maps change, how do they change?

CHAPTER 2

Methodology

Three studies were conducted to investigate the reliability and validity of concept mapping as a practical method to evaluate prior knowledge. The first study explored the feasibility of using a Concept Mapping procedure for further research (Coleman & Sherman, 1992a and later elaborated upon in Coleman & Sherman 1992b). In this study, graduate students in a counseling course (novices) and practicing counselors (experts) were asked to use the pattern noting procedure to show their understanding of a counseling problem. The results of this study indicated subjects could produce a Concept Map (pattern note) that provided information about what subjects knew, Concept Mapping (pattern noting) procedures may be sensitive to differences between experts and novices, and Concept Mapping (pattern noting) may contribute valuable data regarding individual subject's cognitive structure.

Results also indicated a scoring method was needed to conduct the appropriate reliability and validity tests on a Concept Mapping procedure. Furthermore, the pattern noting technique could not be easily scored using a scoring algorithm because of its free-form nature; this technique was more suited to qualitative evaluation. In addition, subjects needed clear instructions on how to do a Concept Mapping procedure and needed training on following the procedure. Although most subjects' maps indicated they could follow pattern noting instructions successfully, there were maps that indicated the opposite. Finally, Concept Mapping procedures did not work well with ill-structured content, a more defined content area to map was required. The counseling content area was somewhat ill structured, which made this subject area difficult for subjects to map.

From this study, it was evident subsequent studies to estimate the reliability and validity of Concept Mapping would require a Concept Mapping procedure that produced a map that lent itself to scoring, that had a proven scoring procedure for the map, that included a training session for subjects, and that defined a well-structured target content area.

The next study (referred to as Study I in this document) was conducted to see if the concept mapping procedure, suggested training program, and scoring algorithm (as described by Novak & Gowin, 1983) were adequate for estimating the reliability and validity of Concept Mapping. Graduate students in an educational research class were trained to do the concept mapping procedure and then were asked to complete a practice concept mapping exercise on the University Library. In the concept mapping investigation, subjects completed a pre- and post-instruction concept mapping and multiple-choice test on regression. The results indicated changes in the concept mapping training session and concept mapping scoring procedure were required before reliability and validity tests could begin. Also, changes in the data collection procedures appeared necessary to increase the number of subjects participating in the study. Study I is fully described in this document.

The last study (referred to as Study II in this document) incorporated the lessons learned from the first two studies. Students in an industrial and systems engineering (ISE) course were trained to do concept mapping, practiced concept mapping by creating a concept map using the University Library as the target concept, and then completed a pre- and post- instruction concept map on measurement in ISE. These data were used to examine the reliability and validity of concept mapping as a tool to assess knowledge prior to instruction. Study II is fully described in this document.

Subjects

Study I

Subjects in Study I were students enrolled in either of two graduate-level statistics sections of the same class at a large land-grant university in Southwest Virginia. Both sections focused on the same content, statistics relevant to the field of education, but were taught by different instructors. Of the 60 students who received data collection materials, five completed all materials. Of these five students, two were pursuing doctoral degrees in education administration, two were pursuing doctoral degrees in special education administration, and one was pursuing a doctoral degree in counseling. The average age of these students was 39, and there were three males and two females. The expert was a statistics instructor who held a Ph.D. degree in Educational Research.

Study II

Students in Study II were undergraduate and graduate students enrolled in an introductory ISE course taught at the same university. There was one instructor for this class. Of the 20 who received data collection materials, 17 completed all materials. The 17 students included one sophomore, five juniors, one senior, and ten graduate students. There were 16 ISE majors and one mechanical engineering major. Their ages ranged from 17 to 38, (11 were between ages 17 and 27, and six were between age 28 and 38). There were ten male and seven female students. The two experts were the instructor of the ISE course and the instructor's teaching assistant (a Ph.D. candidate in ISE).

Instrumentation

Concept mapping and multiple-choice tests

A concept mapping test based on the concept mapping procedure described by Novak and Gowin (1993) (see Appendix 2) and a multiple-choice test on the same target concept were administered to students in Study I and Study II. These tests were given before and after instruction on the target concept. The course instructors helped determine the target concept for the multiple-choice and concept mapping test. Selection criteria for a target concept were a topic that could be covered in a short time (no more than two class periods) and that could be adequately summarized by no more than 20 concepts. This made the concept mapping exercise more manageable for the students.

For Study I, regression (as taught in the course) was chosen as the target concept. Instructors provided course lecture notes and multiple-choice questions (from a test bank) related to regression so the same concept mapping and multiple-choice tests could be developed for both sections of the class. The multiple-choice and concept mapping tests were reviewed by an expert in statistics for accuracy and clarity.

For Study II, measurement from an ISE perspective (as taught in the course) was chosen as the target concept. The course instructor developed student performance objectives pertaining to the target concept and then developed a multiple-choice test covering the student performance objectives. The instructor also generated a list of concept words representing the target concept for the concept mapping test.

Neither the multiple-choice nor the concept mapping tests were timed and they were not counted as a graded assignment for the course. All were taken closed book.

Questionnaire for Content Validity

A Questionnaire for Content Validity was developed for Study II according to guidelines suggested by Crocker and Algina (1986). (The same questionnaire was not necessary for Study I because the pre- and post-tests were not used). The two ISE experts filled out the questionnaire before the concept mapping and multiple-choice tests were distributed and agreed no changes to the list of concepts to be mapped or the questions on the multiple-choice test were necessary.

Student Questionnaire

A questionnaire to estimate the face validity of the concept mapping and multiple-choice tests was administered to the students. This questionnaire also contained demographic questions.

Expert Interview Questions

The two experts in Study II were interviewed to gather additional data for estimating the face and content validity of the concept mapping exercise.

All instrumentation used for Study I can be found in Appendix 3 and all instrumentation used for Study II can be found in Appendix 4.

<u>Procedures</u>

Study I and Study II were conducted in seven steps.

1. The multiple-choice and concept mapping tests were developed and reviewed by experts for content validity.

2. Students were trained to do the concept mapping procedure in a 30-minute training session (the training protocol can be found in Appendix 5).
3. The students completed a practice concept mapping exercise using the University Library as the target concept.
4. The students took the multiple-choice and concept mapping pretests during the same testing period. To determine if performance on one test affected performance on the other, approximately half of the students took the multiple-choice pretest first and half took the concept mapping pretest first. The testing packets were stacked alternating the order of the tests and then were distributed to students sitting in rows. Students were instructed to identify their pretests by a code number so students doing the multiple-choice pretest before the concept mapping pretest could be asked to do the concept mapping post-test before the multiple-choice post-test and vice versa.
5. The course material targeted by the pretests was taught by the course instructors and the students completed the Student Questionnaire.
6. After instruction on the target concepts was complete, students took the multiple-choice and concept mapping post-tests. Students completed the post-tests in the opposite order of the pretests, as described above.
7. The Study II experts were interviewed. A detailed timeline of these steps can be found in Table 1 and Table 2.

Table 1. Timeline for Study I

Date of activity	Activity
February 14-25	1. Instructional materials and test bank questions regarding the topic of regression were gathered from course instructors. 2. Multiple-choice test questions and concept words for the concept mapping test were developed based on these materials.
March 7-10	An expert in statistics reviewed the multiple-choice and concept mapping tests.
March 22/23*	1. Purpose of Study I reviewed with students and consent forms signed. 2. A 30-minute concept mapping training session was given. 3. Practice concept mapping exercise of the University Library was distributed.
March 29/30*	1. Practice concept mapping exercise of the University Library was collected. 2. Concept mapping and multiple-choice pretests on the target concept (regression) were distributed.
April 5/6*	Both classes were canceled this week.
April 12/13*	1. Concept mapping and multiple-choice pretests were collected. 2. Student Questionnaire was distributed.
April 12/13* - April 19/20*	Regression concepts were covered by the course instructors.
April 19/20*	1. Student Questionnaire was collected. 2. Concept mapping and multiple-choice post-tests were distributed.
April 26/27*	Concept mapping and multiple-choice post-tests were collected.

* Two dates appear because the each class section met on separate days.

Table 2. Timeline for Study II.

Date of activity	Activity
May 16-18	1. Measurement from an ISE perspective was selected as the target concept. 2. Student performance objectives, the multiple-choice test and a list of concept words for the concept mapping test were developed by the course instructor. 3. The Content Validity Questionnaire for the multiple-choice and concept mapping tests were completed by two experts in ISE.
May 23	1. The purpose of Study I was reviewed with the students and the consent forms were signed. 2. A 30-minute concept mapping training session was given. 3. Practice concept mapping exercise of the University Library was completed and collected.
May 24	Concept mapping and multiple-choice pretests on the target concept measurement from an ISE perspective were completed.
May 25	Student Questionnaire was completed and collected.
May 25 - May 27	Measurement from an ISE perspective concepts were covered by the course instructor.
May 31	Concept mapping and multiple-choice post-tests were completed.
June 7	Experts were interviewed.

Because of a low response rate (8%) for the pre- and post-tests in Study I, these data were reviewed to understand why the response rate was low and to improve data collection procedures for Study II. The lessons learned are summarized in the next section. The response rate for the University Library practice exercise was higher than that of the pre- and post-tests (32%) in Study I and these data were used to analyze the utility of the scoring procedure. The results of this analysis are summarized in the next section.

Improvements to data collection procedures for Study II

The following lessons learned from a review of data collection procedures in Study I were applied to Study II. The first lesson learned was students needed structured time to complete the data collection materials. Students in Study I were asked to complete data collection materials at home and return them in class the following week (this class met only once a week). Many students failed to turn in the data collection materials. When collecting the data for Study II, the instructor let students complete all data collection materials during class, ensuring students would have time to complete all materials. Second, several students in Study I reported they did not do the concept mapping exercise because they simply did not know enough about regression. Training for Study II stressed that students were not expected to completely understand the meaning of all the concepts they were asked to map. The value of the mapping exercise was in the students' demonstration of what they knew, no matter how basic their understanding. In conjunction, the purpose of pretesting to gather information for *planning* instruction as opposed to *assigning* grades was also stressed during concept mapping training for Study II. It was hoped that de-emphasizing grades would motivate students to complete the concept maps no matter how little they knew about the target concept. Third, the importance of using linking words to describe the relationship between two concepts was reinforced during the concept mapping training for Study II. Several students in Study I submitted concept maps that failed to show the relationships between concepts.

Analysis of the scoring procedure

The fourth lesson learned from Study I was the concept map scoring procedure required modifications and clarifications to increase interrator reliability. Before the concept mapping and multiple-choice tests were scored, two scorers were trained to use the concept mapping scoring method recommended by Novak and Gowin (1993). They practiced with examples provided by Novak and Gowin. Then, using the practice University Library concept maps from Study I, the scorers independently applied the scoring procedure. The correlation between the scores of the two scorers provided the interrator reliability coefficient. The interrator reliability coefficient was .74.

To increase the interrator reliability, the scorers discussed procedures to rate each map and discovered inconsistencies in how the scorers applied the scoring procedure. A list of clarifications was developed and incorporated into the scoring procedure. Scorers again independently scored the practice University Library maps from Study II, using the new scoring criteria. The subsequent interrator reliability coefficient was .85 when the modified scoring procedure was used. See Appendix 1 for the modified scoring procedure.

Improvements to the scoring procedure.

There were several modifications made to the scoring procedure as a result of the review process described above. To clarify scoring of relationships, the term "proposition" was deleted from the scoring instructions and replaced with "relationship." Then, the modified instructions directed scorers to award one point for a relationship unless the relationship showed a high level of understanding between the two concepts, in which case the scorer awarded two points for the relationship. This was to differentiate between the quality of relationships observed in the exercise concept maps from Study I. With the original scoring procedure, the scorers observed some relationships that were stronger or more advanced than others but they had no way to distinguish between them.

The number of points that could be assigned to a crosslink was also changed. In the scoring instructions suggested by Novak and Gowin (1993), a crosslink was assigned a score of ten if it was both correct and significant, and two points if it was correct but not very complex. Rarely did the scorers of the exercise concept maps from Study I find a relationship that was worth ten points; however there were relationships worth more than those scored with two points. Therefore, the modified scoring instructions directed scorers to assign two points to a correct crosslink unless the crosslink showed a high level of understanding between the two crosslinked concepts. In this case a score of four could be assigned.

The Novak and Gowin (1983) scoring instructions did not provide guidance on which branch of the map to use for determining how many levels of hierarchy should be credited to the map. The scoring procedure was clarified by telling scorers to find the longest branch on the map and to count the hierarchical levels within that branch. This number was used to calculate the hierarchical score.

In many cases, the practice maps from Study II were missing a description of the relationship between two concepts. Although the relationship would receive no points, this presented problems in counting hierarchical levels. The modified instructions directed scorers to discount those hierarchical levels without linking words between the concepts. For example, if a branch of concepts with five hierarchical levels contained a segment lacking a description of the relationship between two concepts, then the map would get credit for only four hierarchical levels.

Many of the practice maps from Study I contained arrows showing the flow of the relationships. Because the concept mapping instructions specifically directed students to organize their maps from most general to most specific, the modified instructions directed scorers to ignore the direction of the arrows for a relationship, hierarchy, or crosslink. One reason for this approach was the directional relationship between concepts was not important in scoring the relationship; the stated relationship was being scored. Crosslinks were similar to relationships; the direction of the relationship was not important, but the stated relationship was. Another reason for this approach was the instructions specifically asked students to organize maps from most general to most specific. If the maps were not constructed that way (regardless of the arrows), the maps would not receive points for the hierarchy.

In some cases, students started a branch in a practice map from Study I using the concepts in one context, and then changed the context of the concepts in another branch of the map. For example, "study" might have been used as a verb in terms of "the library is a place to study." However, in the same map, "study" might also have been used as a noun, as in "a study is being conducted" (as shown in Figure 6). The modified instruction directed scorers to ignore this change in context for scoring purposes.

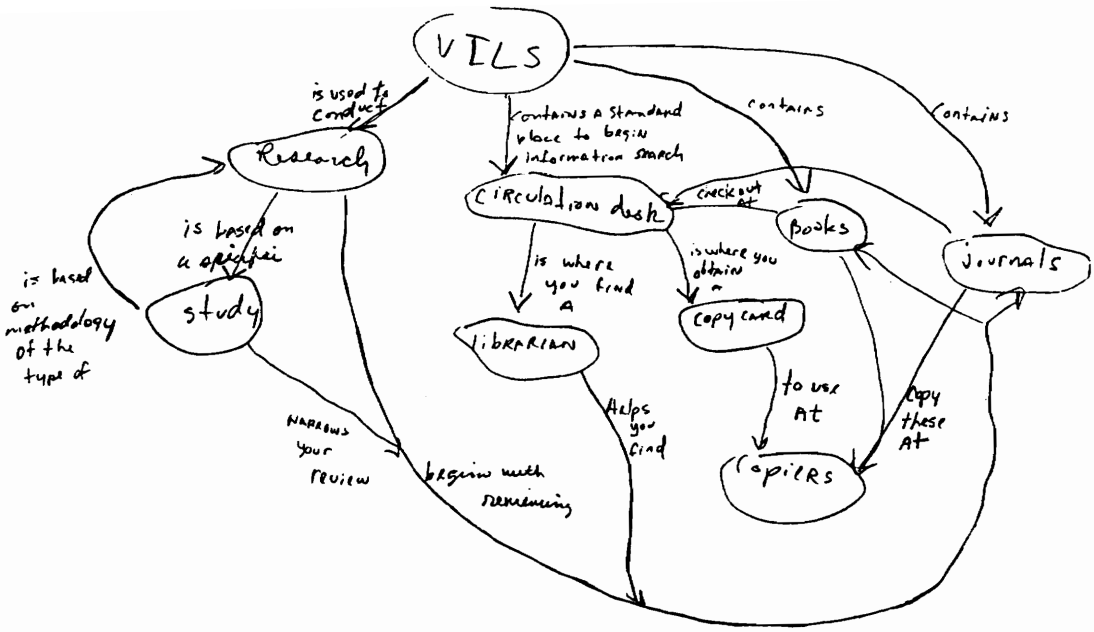


Figure 6. Example - How a student used the concept word "study" in two contexts.

Certain structures in the practice concept maps from Study I presented some problems in scoring. For example, a map would often have concepts arranged in a diamond pattern, with the branch continuing under the bottom point of the diamond (as shown in Figure 7). This made it difficult to tell if a crosslink was being shown. The modified instructions directed scorers to choose one side of the diamond to score as a branch. The other side of the diamond was then considered a crosslink, that is, the two sides connecting with the point of the diamond were considered crosslinks to the main branch made up of the other side of the diamond.

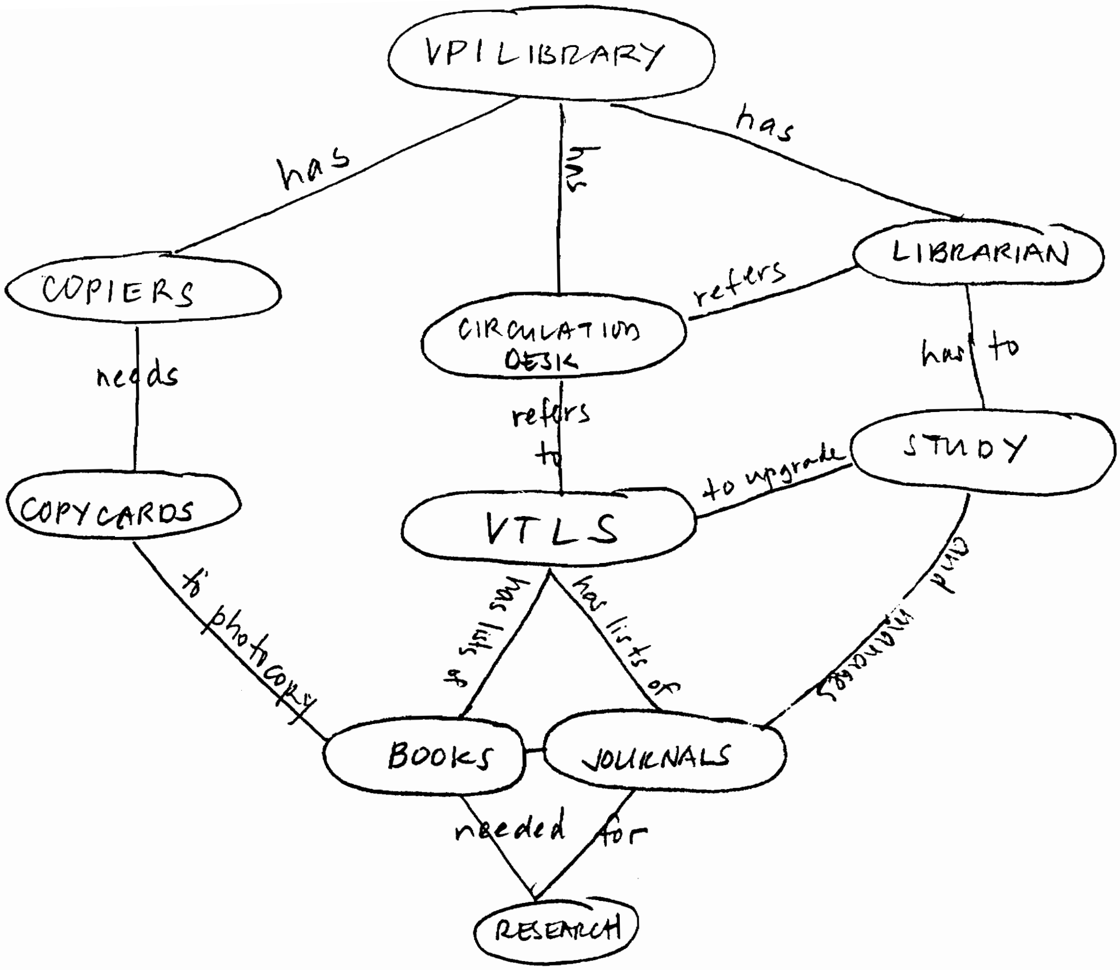


Figure 7. Example - A diamond pattern within a student's concept map.

Another common structure that was difficult to score was a multi-branched node (much like a wagon spoke) without a relationship described on each branch (as shown in Figure 8). In this case only labeled spokes were scored, and if the spoke was part of the branch and was not labeled with a relationship, then it was not counted in the hierarchy. In some cases, a relationship was written across several segments connecting concepts. For this kind of structure the scorer was directed to decide whether the student meant to indicate that the same relationship existed between one concept and several others or if the student simply did not know what kind of relationship existed between the concepts.

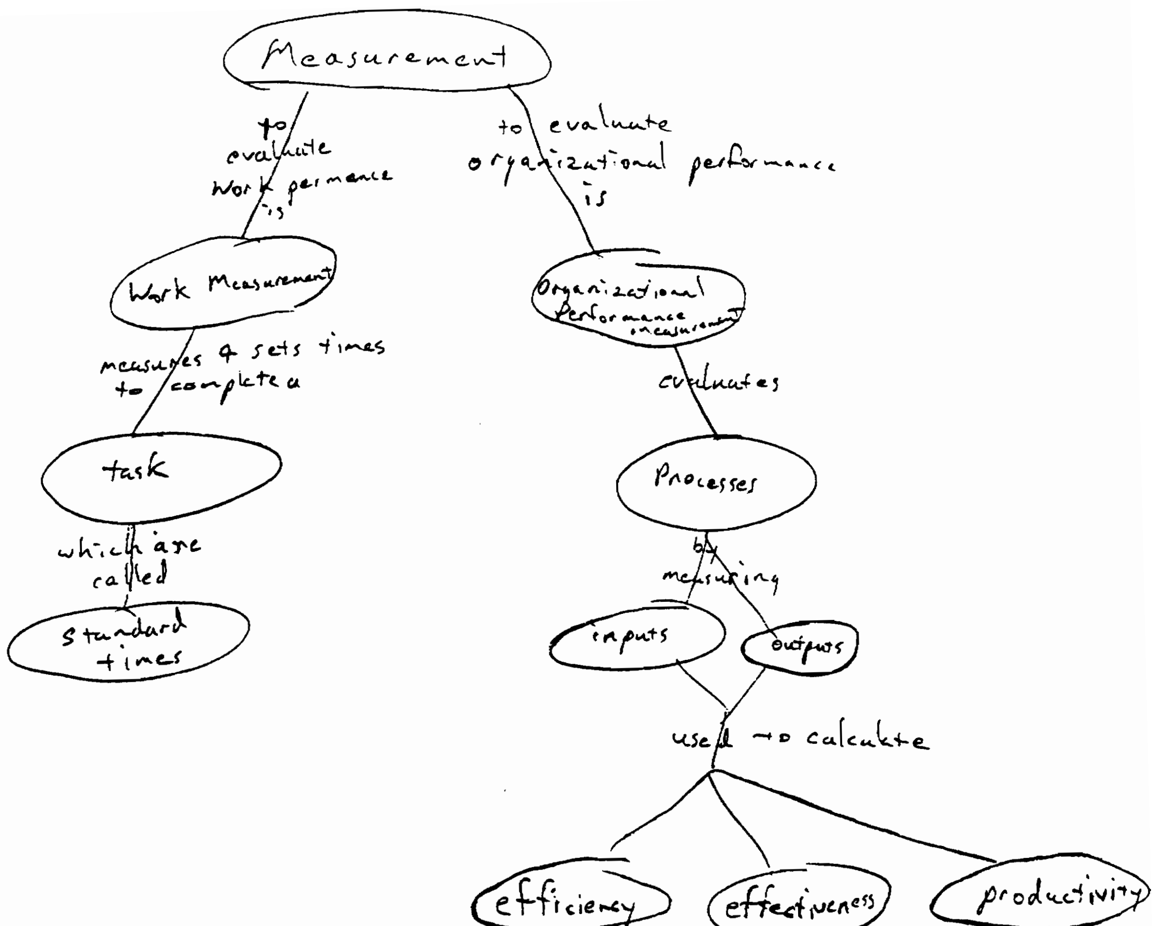


Figure 8. Example - A "wagon spoke" pattern within a student's concept map.

Other problems associated with the Novak and Gowin scoring procedure were not solved. For example, a linear map with few branches but many hierarchies could score higher than a map with few hierarchies, but many branches. This may or may not be desirable depending on the quality of the relationships stated in the map and the content area being mapped. Also, students who understood the scoring algorithm and wanted to maximize their map score, could use this to their advantage. This was not a concern for Study I or Study II because students were not told about the scoring procedure that would be used. But if this scoring procedure is to be used in future studies, this issue needs to be resolved.

Data Analysis

All calculations for the data analyses were done using SPSS on a mainframe or Minitab on an IBM PC. The procedures for estimating reliability and validity are summarized in Table 3, found at the end of this chapter.

Interrator reliability

Interrator reliability is the extent to which two or more scorers agree (Sax, 1989). The higher the degree of agreement between scorers, the higher the interrator reliability. When agreement is high, we can be more confident that variance in students' scores is not due to variance in the scorers but is due to variance in the students' ability being measured.

For this study, the interrator reliability was calculated using the Pearson Product Moment Coefficient (Pearson *r*) on both the concept mapping and multiple-choice pretest scores obtained from two independent scorers. In addition, the average disagreement between the scorers was calculated as another index of interrator agreement. The average disagreement was calculated by taking the absolute value of the differences between the scores of each scorer and finding the mean of the discrepancy scores. An average disagreement of 0 indicated there was no disagreement among scorers. The higher the average disagreement, the more disagreement there was among scorers. The average disagreement was calculated with both the concept mapping and multiple-choice pretests.

Internal consistency

Internal consistency is the degree to which the items on a test are correlated (Brown, 1983). In other words, do all items on the test measure the same trait? If the test is internally consistent, then any one item should be equivalent to any other item in terms of what trait or characteristic it measures. Without internal consistency, it is not clear that variance in student scores is attributable to true variance in ability or is attributable to test items measuring different traits or characteristics. For this study, a Cronbach's alpha coefficient using the concept mapping pretest scores and a KR_{20} coefficient using the multiple-choice pretest scores were calculated to estimate internal consistency.

An item analysis reveals items inconsistent with the total test with regard to students' scores. Performance on one test item is compared to performance on all test items, across all test-takers, to identify those items which correlate poorly to overall test performance (Tuckman, 1985).

Item analysis is usually reported by two coefficients, item discrimination (*D*) and item difficulty (*p*). Item discrimination indicates whether the item differentiates among test takers having varying degrees of ability (Mehrens & Lehman, 1989). In other words, test takers doing well overall on a test should get certain items correct and test takers doing poorly overall on a test should get these same items incorrect. Item difficulty indicates the percentage of test takers answering an item correctly. Items which do not discriminate, and items which no one or everyone answers correctly, weaken the reliability of the test and therefore should be discarded. For this study, an item analysis assisted in interpreting the reliability of the concept mapping instrument.

An item analysis requires test items to be scored separately. Identifying test items is easily accomplished for the multiple-choice test, but it is not as easy for the concept maps. To identify items for the concept mapping test, one of the ISE experts was asked to produce a concept map for "measurement from an ISE perspective." From this map, connected concepts were identified as items. For example, the concept word "work measurement" was connected to the word "standard time." If this relationship was shown on students' maps and the relationship between the concepts was stated correctly, the map received a point. Fourteen items were identified on the expert map. The students' pretest maps were scored for these fourteen items.

Content validity

Content validity provides evidence to draw inferences about test takers' performance to a larger performance domain from which the test items are sampled (Brown, 1983). To draw such inferences, test users must be sure the test items represent the domain about which inferences are to be made. Content validity addresses how well the items represent a sample from the universe of items covering the target domain, and more specifically, "how well the behavior and subject matter called for in the items correspond to the behavior and subject matter identified in the specific objective" (Sax, 1989 p. 294). Content validity is most often judged qualitatively by experts in the target domain (Brown, 1983; Crocker & Algina, 1986; Kerlinger, 1973; Sax, 1989). However, there is quantitative information that may also contribute to estimating content validity. A systematic process for collecting quantitative and qualitative data to estimate content validity is explained in Crocker and Algina (1986) and was used for this study.

First, student performance objectives were obtained from the instructor for the target concept — measurement from an ISE perspective. These objectives defined the content domain for the experts. Second, two ISE experts were surveyed and asked to identify any objectives that went beyond the domain of the target concept, to add any objectives to more fully represent the target concept, and to give an overall assessment of how well the multiple-choice and concept mapping tests emphasized subcategorizes of the target concept in relation to how these subconcepts are emphasized by the objectives. And third, the experts were asked to rate how each multiple-choice question and each concept mapping item matched the student performance objectives in four categories. These categories were: how well the content of the item matched the content of the objective; how well the level of performance complexity required by the item matched the objective; how well the behavior required to answer the question correctly matched the behavior required by the objective (e.g., students learned to add in a column but the test had addition questions written in a row), and; how well the nature of the response matched the objective (e.g., an oral exam requires different response skills than a written exam). To rate each item, the experts used a scale of 1 to 4 corresponding to poor match, fair match, good match, and excellent match.

To gather additional qualitative data to estimate content validity, the two ISE experts were interviewed. The questionnaire and interview questions used to collect the quantitative and qualitative data are found in Appendix 4.

Face validity

A close relative of content validity is face validity. Crocker and Algina (1986) define face validity as the extent to which items appear to measure a construct that it is purported to measure. Face validity is established by evaluating the test by the test takers and considers rather superficial aspects of the instrument. In contrast, content validity is established by a thorough and systematic evaluation of the test by qualified judges and it considers subtle and obvious aspects of relevance. Although face validity may seem an insignificant component of a measure's overall validity, Brown (1983) claims face validity is worth considering because it may motivate examinees to perform at their best because the test appears to measure a meaningful construct.

To determine face validity, students were given a questionnaire concerning their assessment of the face validity of the concept mapping and multiple-choice pretests. See Appendix 3 and 4 for the questionnaire used in Study I and Study II, respectively.

Criterion-related validity (concurrent validity)

Criterion-related validity is estimated by correlating test scores to some outside criterion measure (Sax, 1986). Criterion-related validity is reported with a predictive validity coefficient or a concurrent validity coefficient. Predictive validity coefficients are used when predicting future performance on a criterion measure from present test performance (e.g., predicting performance in the first semester of the college freshman year from SAT scores). Concurrent validity coefficients are used to determine how well performance on one instrument is correlated with performance on another instrument given at the same time and measuring the same trait or ability. Concurrent validity is appropriate to examine when considering substituting one instrument for another (Brown, 1983; Crocker & Algina, 1986; Sax, 1986). However, Brown (1983) suggests substitutions should be made only if there is a strong correlation between the two tests and if the substitute test is more efficient or less expensive than the original test.

For this study, a concurrent validity approach was used to examine the criterion-related validity of the inferences made from the concept mapping scores. The multiple-choice pretest was used as an external criterion measure for knowledge of measurement from an ISE perspective. A Pearson *r* correlation was calculated using the scores from the multiple-choice and concept mapping pretests. The squared validity coefficient indicated the proportion of variance in multiple-choice pretest scores predictable from concept map pretest scores.

In addition, an expert in ISE (the Introduction to ISE course instructor) was asked to rank the concept mapping pretests and post-tests according to how well the students understood the target concept. This ranking was also used as an external criterion measure for knowledge of measurement from an ISE perspective. A Spearman rho correlation was calculated using the rankings resulting from scoring the concept mapping pretests and post-tests and the rankings resulting from the expert evaluation of the concept mapping pretests and post-tests, respectively. The squared validity coefficient indicated the proportion of variance in concept mapping ranks by an expert predictable from concept mapping scores.

Construct Validity

Construct validity is the degree to which we can be sure the test measures unobservable variables (e.g., knowledge of measurement from an ISE perspective) on the basis of observable variables (e.g., performance on a concept mapping test) (Crocker & Algina, 1986; Pedhazur & Schmelkin, 1991). Evidence for construct validity is usually gathered through an examination of hypothesized relationships based on the theory of the nature of the construct being measured. "Performance on the test will vary as the theory underlying the construct would predict" (Mehrens & Lehman, 1991). If the hypotheses are accepted, then it can be concluded the test is useful for measuring the target construct (Crocker & Algina, 1986).

For this study, the construct measured was knowledge of measurement from an ISE perspective. The concept map and multiple-choice tests were used as the criterion measures. If the concept map and multiple-choice tests measured knowledge of measurement from an ISE perspective, then there should have been a significant increase between the pre- and post- test scores for both tests. Because the concept map and multiple-choice tests were administered together, it was possible that the first test taken affected performance on the second test. Therefore, two 2x2 repeated measures designs were set up to analyze the concept map and the multiple-choice pretest scores.

The two independent variables in the 2x2 repeated measures designs were the time of the test (pre or post) and the order in which students took the test (first or second). The dependent variable for the first repeated measures was the concept mapping test scores and, for the second, the multiple-choice test scores. Three questions were investigated with this procedure: Is there a significant difference between scores on the pre- and post-test?; Is there a significant difference between scores when the multiple-choice test is taken first and when the concept map test is taken first?, and; Is there interaction between the time of the test and the order in which the tests are taken?

Mapping Ability

Students' ability to do the concept mapping procedure may have been a variable affecting performance on the concept mapping test, and therefore may have affected reliability. Variance in the scores could then have been attributed to students' mapping ability rather than differences in students' understanding of the target concept. To determine if students in Study II were mapping at similar levels of ability, students were asked to map a target concept that all students should know quite well, the University Library. Any variance in map scores was attributed mainly to mapping skill, not knowledge of the library. In addition to examining the distribution of map scores for the practice exercise, a qualitative analysis of the maps was done to further describe the differences in the students' maps and to estimate the difference in students' mapping abilities.

Table 3. Procedures for estimating reliability and validity.

	Major error source	Statistical treatment of data	Data collection procedure
RELIABILITY			
Interrator reliability	To demonstrate that variance in students' test scores is *not* primarily due to differences in the way the two scorers scored the tests.	Pearson Product Moment Coefficient (Pearson *r*) and the average absolute disagreement were calculated using the concept mapping and multiple-choice pretest scores from the two scorers.	Two scorers scored the concept mapping and multiple-choice pretests.
Internal Consistency	To demonstrate that variance in students' tests scores is *not* primarily due to inconsistency in content sampling (i.e., the test items are measuring the same construct).	Cronbach's alpha using the concept mapping pre- and post-test scores and KR_{20} using the multiple-choice pre- and post-test scores were calculated. Item analysis was completed	Administered the concept mapping and multiple-choice pre- and post-tests.
VALIDITY			
Content validity	To demonstrate the items on the concept map and multiple-choice tests are representative of the universe of items relevant to this topic.	1. Qualitative - The opinions of two experts were analyzed. 2. Quantitative - Expert ratings of each concept mapping and multiple-choice test item on four different content-related characteristics were averaged.	1. Experts were interviewed. 2. Experts completed the Questionnaire for Content Validity.

Table 3 (continued)

	Major error source	Statistical treatment of data	Data collection procedure
Face validity	To determine if subjects feel the concept mapping and multiple-choice tests measure what they are supposed to measure. And to determine how good a measure subjects believe concept mapping and multiple tests are.	Qualitative - The opinions of students and experts were analyzed.	Questions regarding the face validity of concept mapping and the multiple-choice tests were included on the Student Survey and were asked during the expert interviews.
Criterion-related validity (concurrent validity)	To determine if performance on the multiple-choice test can be predicted by performance on the concept mapping test. To determine if performance on the concept mapping test can be predicted by the results of an expert sorting the concept maps by the degree to which the student demonstrated understanding of the target topic.	Pearson *r* using the scores of the multiple-choice and concept mapping tests was calculated. Spearman rho using concept mapping scores and results from ranking the maps was calculated.	Adminstered the concept mapping and mulitple-choice pre- and post-tests. An expert ranked the concept maps from the pre- and post-test.
Construct validity	To determine if the concept maps are measuring prior knowledge - a change in the scores of the concept maps and multiple tests (i.e., higher scores) after instruction was expected.	A repeated measures analysis was conducted.	Administered the concept mapping and multiple-choice pre- and post-tests.

CHAPTER 3

Results

This chapter contains the results of the reliability and validity analyses from the Study II data. Study I yielded so few usable data they were not amenable to the reliability and validity analyses conducted for Study II. Study I data were used primarily to improve the scoring and data collection procedures and the data collection materials for Study II. Study I data are referred to in this chapter only to compare the interrator reliability of the concept mapping exercise of Study I with the concept mapping pretest of Study II so that differences may be addressed.

The results presented in this chapter are organized by reliability and validity analyses. Each section may report qualitative and/or quantitative analyses. Discussion of the results follows each section to explain them in the context of the individual reliability or validity analysis and in the context of the study as a whole. Results are summarized in Table 4.

Interrator reliability

The interrator reliability correlation between the two independent scorers scoring the multiple-choice pretests was 1.0. The average disagreement between scores was 0. The interrator reliability correlation between two independent scorers scoring the concept maps using the revised Novak and Gowin (1993) scoring procedure was .67. The average disagreement between scorers scoring the concept mapping tests was 4.21.

Table 4. Summary of results from the data analysis.

	Major error source	Statistical treatment of data	Results Summary	
RELIABILITY			concept mapping tests	multiple-choice tests
Interrator reliability	To demonstrate that variance in students' test scores is *not* primarily due to differences in the way the two scorers scored the tests.	Interrator reliability coefficient Average disagreement	.67[1] .76[2] 4.21[1] .63[2]	1.0 0.0
Internal Consistency	To demonstrate that variance in students' test scores is *not* primarily due to inconsistency in content sampling (i.e., the test items are measuring the same construct).	Reliability coefficients pretest post-test	.3141[2] .3830[2]	.2320 .0070 .3070[3]
VALIDITY				
Content validity	To demonstrate the items on the concept map and multiple-choice tests are representative of the universe of items relevant to this topic.	Qualitative Quantitative (avg. item rating from a 4-pt scale)	very positive 3.725	very positive 3.85

[1] Concept mapping tests scored using the modified Novak and Gowin scoring method.
[2] Concept mapping tests scored using the new scoring method.
[3] Multiple-choice post-test plus unit test questions.

Table 4 (continued)

	Major error source	Statistical treatment of data	Results Summary concept mapping tests	multiple-choice tests
VALIDITY (con.)				
Face validity	To determine if subjects feel the concept mapping and multiple-choice tests measure what they are supposed to measure. And to determine how good a measure subjects believe concept mapping and multiple-choice tests are.	Qualitative analysis	Approximately 50% of students preferred concept mapping tests, however all expressed low confidence in the measure. Experts responded favorably.	Approximately 50% of students preferred multiple-choice tests, however all expressed low confidence in the measure. Experts responded favorably.
Criterion-related validity (concurrent validity)	To determine if performance on the multiple-choice test can be predicted by performance on the concept mapping test.	Concurrent validity coefficients		
		Concept mapping pretest[1]		-0.219
		Concept mapping pretest[2]		0.210
		Concept mapping pretest ranking	-0.501[1]	
			-0.544[2]	
	To determine if performance on the concept mapping test can be predicted by the results of an expert sorting the concept maps by the degree to which the students showed understanding of the target topic.	Concept mapping post-test[1]		0.176
		Concept mapping post-test[2]		0.082
		Concept mapping post-test ranking	-0.629[1] -0.197[2]	
		Concept mapping post-test[1]		0.005[3]
		Concept mapping post-test[2]		0.128[3]

[1] Concept mapping tests scored using the modified Novak and Gowin scoring method.
[2] Concept mapping tests scored using the new scoring method.
[3] Multiple-choice post-test plus unit test questions.

Table 4 (continued)

	Major error source	Statistical treatment of data	Results Summary	
VALIDITY (con.)			concept mapping tests	multiple-choice tests
Construct validity	To determine if the concept maps are measuring prior knowledge - a change in the scores of the concept maps and multiple-choice tests (i.e., higher scores) after instruction indicates the device measured a change that was expected.	Repeated measures by order and time of the test.		
		Time	n.s.[1] significant[2]	significant
		Order	n.s.[1,2]	n.s.
		Time x Order	n.s.[1,2]	n.s.

[1] Concept mapping tests scored using the modified Novak and Gowin scoring method.

[2] Concept mapping tests scored using the new scoring method.

[3] Multiple-choice post-test plus unit test questions.

The interrator reliability was lower for the concept mapping pretests for Study II (.67) as compared to the concept mapping exercise maps for Study I (.85, as reported in Chapter 2). The difference may be because the scorers who scored the concept mapping pretests in Study II were less familiar with the scoring procedure and did not have the opportunity to discuss scoring decisions as did the concept map exercise scorers in Study I. For future studies, the scoring procedures need to be clarified so novice scorers can reliably score concept maps.

The average disagreement for the concept mapping tests could have ranged from 0 to 18, 18 being the largest discrepancy between the scores assigned to one test. The relatively low average disagreement (4.21) is consistent with the interrator reliability coefficient (.67). While the interrator reliability coefficient indicates how consistently the scorers ranked the students taking the tests, it does not precisely indicate how much the scorers disagreed For example, one scorer could consistently assign low scores to a set of tests and the other scorer consistently assign high scores to a set of tests. As long as the tests were ranked relatively the same according to the assigned score, the interrator reliability would be somewhat high. The average disagreement for the concept mapping test showed the interrator reliability was high and the difference between scores was high as well. The average disagreement enriches the interpretation of the interrator reliability of an instrument.

Because a new scoring method was developed for the concept mapping test for an item analysis, an interrator reliability was calculated between two independent scorers using the new scoring method on the concept mapping pretests. The interrator reliability coefficient was .76 and the average disagreement was .63 (disagreement ranged from 0-4 points). This scoring method required less subjectivity on the part of the scorers, perhaps accounting for the higher interrator reliability coefficient and lower average disagreement.

Internal consistency

Although this study examined the reliability and validity of concept mapping in the context of a pre-learning assessment, the reliability of the pre- and post-tests was examined to provide a more complete picture of the reliability of the concept mapping and multiple-choice tests. Item difficulty, standard deviation, and item discrimination for each item of each test are reported in Tables 5-9. The internal consistency coefficient for each test is summarized in Table 10.

Table 5. Item difficulty and item discrimination for the concept mapping pretest, scored with the new scoring method.

Item	Item Difficulty	Standard Deviation	Item Discrimination
1	.3333	.4851	.3057
2	.3889	.5016	.3661
3	.6667	.4851	.2822
4	.4444	.5113	-0.3308
5	.1111	.3234	.2599
6	.1111	.3234	.2599
7	.2222	.4278	-0.3199
8	.5000	.5145	-0.2176
9	.3333	.4851	.3057
10	.2778	.4609	.1850
11	.4444	.5113	.2986
12	.3333	.4851	-0.0202
13	.1667	.3835	.2668
14	.3889	.5016	.0397
Average	.3373		

Table 6. Item difficulty and item discrimination for the concept mapping post-test, scored with the new scoring method.

Item	Item Difficulty	Standard Deviation	Item Discrimination
1	.4444	.5113	.1383
2	.3889	.5016	.1374
3	.7222	.4609	.3618
4	.2778	.4609	.3044
5	.1111	.3234	-0.2007
6	.0556	.2357	.2850
7	.2778	.4609	.1638
8	.5556	.5113	.0450
9	.7778	.4278	.4467
10	.6111	.5016	.2590
11	.6667	.4851	.1670
12	.7222	.4609	.5167
13	.2778	.4609	-0.3766
14	.8889	.3234	-0.4331
Average	.4841		

Table 7. Item difficulty and item discrimination for the multiple-choice pretest.

Item	Item Difficulty	Standard Deviation	Item Discrimination
1	.3889	.5016	.0220
2	.0000	.0000	.0000
3	.4444	.5113	.0550
4	.1111	.3234	-0.1097
5	.1111	.3234	-0.1097
6	.1667	.3835	.4243
7	.5556	.5113	.0550
8	.7222	.4609	.3478
Average	.3571		

Table 8. Item difficulty and item discrimination for the multiple-choice post-test.

Item	Item Difficulty	Standard Deviation	Item Discrimination
1	.7778	.4278	-0.2703
2	.1111	.3234	.0413
3	.9444	.2357	-0.3287
4	.8889	.3234	.2426
5	.3333	.4851	.1414
6	1.0000	.0000	.0000
7	.6111	.5016	.1722
8	.8333	.3835	.0000
Average	.6429		

Table 9. Item difficulty and item discrimination for the multiple-choice post-test with questions from the unit test added.

Item	Item Difficulty	Standard Deviation	Item Discrimination
1	.7778	.4278	-0.3372
2	.1111	.3234	.1369
3	.9444	.2357	-0.3123
4	.8889	.3234	.2008
5	.3333	.4851	.1280
6	1.0000	.0000	.0000
7	.6111	.5016	.3568
8	.8333	.3835	.0969
9	.8333	.3835	.3541
10	.8889	.3234	.3521
11	.8889	.3234	.2008
Average	.7111		

Table 10. Reliability coefficients for the pre- and post-tests.

Test	Internal Consistency Coefficient
Concept mapping pretest	.3141
Concept mapping post-test	.3830
Multiple-choice pretest	.2320
Multiple-choice post-test	.0070
Multiple-choice post-test plus unit questions	.3070

The item difficulty indicates the proportion of subjects who answered the item correctly. The higher the item difficulty (*p*), the more subjects answered the item correctly (Sax, 1989). Desirable item difficulties vary according to the purpose of the test (Sax, 1989). If the purpose of the test is to identify those examinees who excel, then low item difficulties are desirable. If the purpose of the test is to identify those examinees who have reached a level of mastery, then high item difficulties are desirable. The average difficulty indicates the overall difficulty of the test.

The item discrimination is the correlation between performance on one item to performance on the test (Brown, 1983). A positive item discrimination coefficient indicates that, overall, subjects answering the item correctly scored high on the test and subjects answering the item incorrectly scored low on the test. A negative correlation indicates that, overall, subjects answering the item correctly scored low on the test and subjects answering the item incorrectly scored high on the test. A correlation of 0 indicates there is no relationship between answering the item correctly and the total score on the test. Desirable items show a positive item discrimination and Brown (1983) suggests an item should have a discrimination correlation of .20 or higher to be included in a test.

The relationship between item difficulty and item discrimination is important to understand the internal consistency coefficients associated with the concept mapping and multiple-choice pretests. Any test with mostly difficult items (low p value) or mostly easy items (high p value) is not able to discriminate among examinees. For example, item two in the multiple-choice pretest (see Table 7) had an item difficulty of 0 (meaning no one answered the item correctly) and an item discrimination of 0 (the item is unable to discriminate among subjects). Item six in the multiple-choice post-test with questions from the unit test added (see Table 9) had an item difficulty level of 1.000 (meaning everyone answered the item correctly) and an item discrimination of 0 (the item was unable to discriminate among students). When a test cannot discriminate among examinees, the item discrimination coefficients are low and the test will have a low internal consistency coefficient, as is the case with the concept mapping and multiple-choice pretests. In both cases, the item difficulties were rather low, indicating the items were difficult for the students. This resulted in poor item discrimination coefficients and poor internal consistency coefficients.

One explanation for the difficulty levels of the pretest items is students had not yet been taught the concepts tested by the concept mapping and multiple-choice pretests. Group homogeneity can affect the reliability coefficient because of the lack of variation among individuals (Crocker & Algina, 1986). As expected, the students were homogeneous in terms of their knowledge of measurement from an ISE perspective — they all knew very little, negatively affecting the internal consistency coefficient.

The average difficulty for the concept mapping post-test, the multiple-choice post-test, and the multiple-choice post-test plus the unit questions increased from the respective pretests, indicating more students answered the questions correctly on the post-tests. (Whether this represents a significant increase in scores from pre- to post-test is addressed in the discussion of construct validity later in this chapter.) This was expected considering students had just received instruction on the concepts targeted in the post-tests. However, the item discrimination indices did not indicate the post-tests could discriminate well among students and the internal consistency coefficients were low. Subjectivity in scoring and test length could explain these results.

Subjectivity in scoring can weaken reliability because of the inconsistencies introduced by subjective scoring methods, creating random error in the scores (Sax, 1989). This may have been the case with the concept mapping post-tests (and the pretests as well). Because of the odd structures encountered in several of the students' maps, the scorers had to interpret what the students meant by their maps. This may have introduced random error into the map scores because the scorers could interpret the maps differently, resulting in a different map score and weakening the reliability.

Although subjectivity of the scoring method was not a factor in the multiple-choice tests, the low number of test items could have affected the reliability of the tests. Crocker and Algina (1986) and Brown (1983) both state longer tests tend to be more reliable than shorter tests. The multiple-choice pretest had eight questions. These questions were directly tied to the student performance objectives, but more questions were required to increase the test's reliability. Because the multiple-choice test was necessary for the concurrent validity analysis, questions were added to it to improve the internal consistency. The instructor had three questions from a unit test directly related to the student performance objectives covered by the multiple-choice test. When these questions were added to the multiple-choice post-test, the internal consistency increased to .3001.

Content validity

Qualitative and quantitative data were collected from two experts to examine the content validity of the concept mapping and multiple-choice tests. Overall, the results of an analysis of both quantitative and qualitative data indicated the concept mapping and multiple-choice tests generally measured what they were supposed to measure, according to the experts in the content area. However, the qualitative data analysis revealed concern regarding the ability of the concept mapping test procedure to measure students' knowledge.

Questionnaire for Content Validity

The quantitative results from the Questionnaire for Content Validity are summarized in Table 11 and 12. The Questionnaire for Content Validity data can be found in Appendix 6.

Table 11. Expert ratings of the concept mapping test items from the Questionnaire for Content Validity (average of responses on a 4-point scale are given)

multiple choice test item number:	subject matter	level of complexity of the performance required	format of presentation	mode of responses
1	3.5	4.0	3.5	4.0
2	3.5	3.5	3.5	4.0
3	4.0	4.0	4.0	4.0
4	4.0	4.0	4.0	4.0
5	2.5	3.5	3.0	4.0
6	3.5	3.5	4.0	4.0
7	2.5	3.0	3.5	4.0
8	4.0	4.0	4.0	4.0
Overall average	3.5	3.7	3.7	4.0

Table 12. Expert ratings of the multiple-choice test items from the Questionnaire for Content Validity
(average of responses on a 4-point scale are given)

concept mapping item	subject matter	level of complexity of the performance required	format of presentation	mode of responses
work measurement	4.0	4.0	4.0	4.0
organizational performance measurement	4.0	4.0	4.0	4.0
organizational system	4.0	3.5	4.0	3.5
efficiency	4.0	4.0	4.0	4.0
effectiveness	4.0	4.0	4.0	4.0
productivity	4.0	4.0	4.0	4.0
standard time	4.0	4.0	4.0	4.0
inputs	4.0	3.5	3.5	4.0
outputs	4.0	3.5	3.5	4.0
processes	3.5	3.5	3.5	4.0
tasks	3.5	4.0	4.0	4.0
Overall average	3.9	3.8	3.8	3.9

When asked about the student performance objectives, experts agreed no objective went beyond the domain of measurement (from an ISE perspective, as presented in the Introduction to ISE course) and no additional objectives were necessary to more fully represent the domain. Additionally, both experts agreed there was a balance between the emphasis on subcategories (i.e., concepts such as efficiency, related to measurement from an ISE perspective) within the objectives and the emphasis on the subcatagories as reflected by test items.

The average rating for the match between the subject matter contained in the objectives and the subject matter contained in the test items was 3.9 for the concept mapping test and 3.5 for the multiple-choice test (see Table 11 and 12), indicating that the test items on the concept mapping and multiple-choice tests were generally on target. Outliers on the multiple-choice test included item 5 which received a rating of 3 (good match) from Expert A and 2 (fair match) from Expert B, and item 7 which received a rating of 2 from Expert A and a 3 from Expert B. Item 5 asked students to identify which one out of five given scenarios was not a measure of productivity and corresponded to objective number 2, define effectiveness, efficiency, and productivity. Item 7 asked the student to identify a situation showing the difference between these two concepts and corresponded to objective number 4, explain the relationship between work measurement and organizational performance.

The average rating for the match between level of complexity of performance required by the objectives and the test items was 3.8 for the concept mapping test and 3.7 for the multiple-choice test. Neither expert rated any item lower than 3 (good match). The average rating for the match between the format of presentation and the objectives was 3.8 for the concept mapping test and 3.7 for the multiple-choice test and neither expert rated any item lower than 3. The average rating for the match between mode of response and the objectives was 4.0 for the multiple-choice test and 3.9 for the concept mapping test. Again, neither expert rated any item lower than 3.

Expert Interview

During the interview, the experts agreed the concept mapping tests measured the concepts they were supposed to measure. However, the experts were not confident the concept mapping procedure could measure what students know. Although they appeared to be on target, the ability of students to create a concept map showing what they knew was questioned by the experts, despite the results of the Questionnaire for Content

Validity where both experts agreed the format of presentation and mode of responses presented no problem with the content validity. After observing students taking the concept mapping test, both experts expressed concern about the mapping abilities of the students as evidenced by one comment, "Concept mapping skills were poor and may have made a difference. If they (the students) had better concept mapping skills, it may have been a more effective exercise."

Another reason experts provided for their lack of confidence in concept mapping to measure students' knowledge is that the concept mapping test was more dependent on student effort than the multiple-choice test. One expert noticed that not all students put their best efforts into completing the concept mapping tests. This expert suggested students knew they would not be graded so they did not do their best work. Students also knew multiple-choice tests would not be graded, but this test required less effort to perform well. The other expert commented, "Multiple-choice tests are not as dependent on student effort." Therefore, the experts were reluctant to express confidence in the ability of the concept mapping test to fairly measure students' knowledge.

Despite concerns about concept mapping as a way to evaluate students, the experts expressed positive comments toward concept mapping as a tool for gathering information for instructional planning. Experts preferred concept mapping over multiple-choice tests for instructional planning because of the richness in the results provided by concept mapping. One expert said, "The multiple-choice test question is either right or wrong, you can't get much more than that, like why students answered the question wrong." and, "With the multiple-choice test, each question is independent, I can't see how they (students) fill in the gaps." The other expert said, "Concept maps showed knowledge structure, how do the pieces fit together - shows [sic] the big picture."

Preparation requirements for a concept mapping test was another plus, according to the experts. One expert commented, "Students can't mimic with concept mapping - they must understand the materials. You can memorize and do well on a multiple-choice test, but not with a concept mapping test." Regarding guessing, both experts liked the fact that there is a very small chance students could guess correctly on a concept mapping test, as opposed to a multiple-choice test where the student has a much greater chance of guessing correctly.

The ease with which the instructor constructed the concept mapping test was another reason for liking it. One expert (the one who constructed the concept mapping and multiple-choice tests) said, "It is easier to construct a concept mapping exercise" and then said writing multiple-choice test questions is time consuming and difficult.

Face validity

The Student Questionnaire from Study II was the data source for estimating face validity (the response rate for Study I was very poor (5%) and these data were discarded). The Student Questionnaire data can be found in Appendix 8.

In general, the results of the questionnaire revealed students felt positive toward the concept mapping test as a way to gather information for instructional planning. When students were asked to name a preference for either the concept mapping test or the multiple-choice test as a way to measure what students know, slightly more than half selected the concept mapping test. However, when asked about their confidence in either test to let them show what they knew, students responded they had little confidence in either test. A detailed analysis of students' responses to the questionnaire follows.

The first question asked students to indicate which they preferred for measuring what they knew, concept mapping or multiple-choice tests, and why. Nine preferred the concept mapping test and eight preferred the multiple-choice test. Those who preferred concept mapping agreed they liked the way concept mapping let them show relationships between concepts. They felt this was more demonstrative of what they knew than their performance on a multiple-choice test. One student responded, "Sometimes you may know what you want to say, but with a multiple choice test you may not find the answer that you are looking for. Concept mapping allows you to state exactly what you think something means." Similarly, the opportunity to explain responses was considered a plus for concept mapping, as opposed to the multiple-choice test which one student described as a "guessing test." Another student said, "Concept mapping allows freedom of expression. Allows logical statements to be presented to back up choices." The downside of concept mapping, as reported by its supporters, was the amount of subjectivity in scoring the results. For example, one student was concerned about concept maps being more subjective than multiple-choice tests.

There were a number of reasons students reported for preferring the multiple-choice test. Two students preferred multiple-choice tests because they are less subjective. One wrote, "The concept map is a little more complicated and there is more than one answer, there are not exact answers." Several students preferred how multiple-choice test items converge on one correct answer, as opposed to a concept mapping test which diverges from one concept, suggesting a variety of correct answers. One student explained, "I prefer multiple-choice tests. Because, generally, there is only one "correct" answer in a multiple-choice question. If I know the answer, it's quite easy to select the correct response. On the concept mapping tests, there are too many ways to present the answer and it's hard to tell what would be considered the correct response." One student preferred multiple-choice tests because there are more cues provided on this kind of test, "I preferred the multiple-choice question [sic] simply because the questions provided clues to help sort my thought [sic] and come up with a correct answer." Finally, some students reported a higher comfort level with the multiple-choice test than with the concept mapping test because they were more familiar with the multiple-choice test format. One student suggested he/she might prefer the concept mapping test after more practice with this method.

When asked to describe how the information provided on the concept mapping test was different from that on the multiple-choice test, students most often stated they were able to describe relationships between concepts which they could not do on the multiple-choice test. One student wrote, "The concept maps show interrelationships between items that could not be covered by a multiple-choice test." Another student commented that, "The concept map showed more of what I knew about the subject because of the relations." Similarly, two students agreed they could give a more complete picture of what they knew through concept mapping. Students reported concept mapping let them "give a relationship between all the terms instead of separating them all into separate questions." In contrast, two other students felt the information provided on the concept mapping test was vague compared to the information provided on the multiple-choice test.

Several students contrasted the concept mapping and multiple-choice test process in describing the differences they perceived in the two measurement methods. Two students said completing the concept mapping test required more thought than completing the multiple-choice test (a comment also made by an expert during the expert interview). One student wrote, "More thinking and organization skills required [sic]. Sometimes it also depends on creative thinking of how to link all concepts on the problems in a cohesive fashion." And two students commented the process for completing the multiple-choice test was easier because multiple-choice tests "may contain key words to trigger memory of meaning."

The third question asked if students felt their knowledge of measurement (from an ISE perspective) was fairly assessed using the concept mapping test. The majority of students replied "no" because they could not describe all they knew about the topic through their concept maps. This is demonstrated in one student's response, "I feel that I could provide a dozen more maps of my knowledge of measurement, depend [sic] on which perspective I want to present it [sic]." In addition, several students commented the concept mapping test would not provide a thorough picture of what they knew about measurement, "I don't think that the concept map gives a good assessment of one's knowledge in any subject. No 1 test can give an [sic] complete assessment of one's knowledge in a subject. The more types of tests the better."

Students also responded "no" to this question because they were not yet comfortable with the process of concept mapping. One student wrote, "..because I am still not sure about mapping. I don't think I could map out something I know very well right now and give as much info [sic] as it was to write an essay." The subjectivity of the maps was also given as a reason for a lack of confidence in concept mapping.

One student responding "yes" to this question was also concerned about the subjectivity in the evaluation of the map, and wrote, "...different people see things in different ways so a different answer isn't necessarily wrong." Another student responded "yes" because concept mapping let him or her show a broad perspective of his or her understanding of the topic. And one student responding "yes" explained, "if I was not knowledgeable it would probably show pretty easily (on the concept map)."

The fourth question asked students if they felt their knowledge of measurement (from an ISE perspective) was fairly assessed using the multiple-choice test. The majority of the students responded "no" to this question. One student suggested multiple-choice tests required more memorization than understanding. Another student thought problem solving and creative processes were not assessed through multiple-choice tests. And another student commented that multiple-choice tests do not provide an overall picture of what someone knows.

Those responding "yes" to this question commented that multiple-choice tests make it clear to the test taker what the scorer will be scoring. One student said the reason he or she had confidence in multiple-choice tests was because sufficient knowledge was required to answer test items correctly. Another student said "yes" even though "it (multiple-choice test) allows a lucky guess to be mistaken for concrete knowledge."

Students overwhelmingly agreed concept mapping could be used by an instructor to assess what students know prior to instruction (with the intention of using the information to plan instruction and not assign grades). One student commented, "It (concept mapping) would provide the instructor with a general level of understanding that the class has mastered to that point." Another student wrote, "It (concept mapping) points out glaring deficiencies, erroneous concepts as well as strengths and weaknesses." A student noted that no one way should be used to determine what students know prior to instruction and concept mapping should be part of a variety of tests. Of those responding that concept mapping was not a good choice for gathering information for instructional planning, one student suggested concept mapping would be better for assessing knowledge rather than planning instruction. Another student was concerned a map could be constructed without a "real idea of the concept."

Criterion-related validity (concurrent validity)

The validity coefficients for the concept mapping tests are summarized in Table 13. Expectancy tables for each comparison can be found in Tables 14-23. The expectancy tables show the percentage of students scoring within a range of concept mapping test scores by their concept mapping ranks or their multiple-choice test scores. Patterns should have emerged if the concept mapping test scores were strongly correlated with the concept mapping ranks or with the multiple-choice test scores. As is evident by Table 13 and Tables 14-23, the only significant correlations were between the concept mapping test scores and the concept mapping rankings. (According to Ferguson & Takane, 1989, with df = 17, and at a level of .05 significance, a correlation of .456 is required for significance.) The remaining correlations indicate the predictability of multiple-choice test scores from concept mapping test scores was practically random.

Table 13. Validity coefficients for the concept mapping tests.

	Multiple-choice pretest scores
Concept mapping pretest scores (using the modified Novak and Gowin scoring method)	-0.219
Concept mapping pretest scores (using the new scoring method)	0.210
	Multiple choice post-test scores
Concept mapping post-test scores (using the modified Novak and Gowin scoring method	0.176
Concept mapping post-test scores (using the new scoring method)	0.082
	Multiple-choice post test plus unit questions scores
Concept mapping post-test scores (using the modified Novak and Gowin scoring method)	0.005
Concept mapping post-test scores (using the new scoring method)	0.128
	Concept mapping test rankings
Concept mapping pretest scores (using the modified Novak and Gowin scoring method)	-0.501*
Concept mapping pretest scores (using the new scoring method)	-0.544*
Concept mapping post-test scores (using the modified Novak and Gowin scoring method)	-0.629*
Concept mapping post-test scores (using the new scoring method)	-0.197

*$p \leq 0.05$

Table 14. Expectancy table showing percentages of students in each concept mapping pretest score range (scored using the modified Novak and Gowin scoring method) who received various multiple-choice pretest scores.

Concept mapping pretest score ranges	Multiple-choice pretest scores								
	0	1	2	3	4	5	6	7	8
16-20	0	0	0	0	5.5	0	0	0	0
21-25	0	0	0	5.5	11.1	0	0	0	0
26-30	5.5	11.1	5.5	0	0	5.5	0	0	0
31-35	0	11.1	11.1	11.1	0	0	0	0	0
36-40	0	0	5.5	5.5	0	0	0	0	0
41-45	0	0	0	5.5	0	0	0	0	0

Table 15. Expectancy table showing percentages of students in each concept mapping pretest score range (scored using the new scoring method) who received various multiple-choice pretest scores.

Concept mapping pretest score ranges	Multiple-choice pretest scores								
	0	1	2	3	4	5	6	7	8
0-2	5.5	0	5.5	0	5.5	0	0	0	0
3-5	0	11.1	5.5	11.1	5.5	5.5	0	0	0
6-8	0	5.5	16.7	16.7	5.5	0	0	0	0
9-11	0	0	0	0	0	0	0	0	0
12-14	0	0	0	0	0	0	0	0	0

Table 16. Expectancy table showing percentages of students in each concept mapping post-test score range (scored using the modified Novak and Gowin scoring method) who received various multiple-choice post-test scores.

Concept mapping post-test score ranges	Multiple-choice post-test scores								
	0	1	2	3	4	5	6	7	8
16-20	0	0	0	0	0	0	5.5	0	0
21-25	0	0	0	0	5.5	0	5.5	0	0
26-30	0	0	0	0	5.5	5.5	5.5	5.5	0
31-35	0	0	0	0	5.5	16.7	5.5	0	0
36-40	0	0	0	0	0	5.5	0	11.1	0
41-45	0	0	0	0	0	5.5	0	0	0
46-50	0	0	0	0	0	5.5	0	5.5	0

Table 17. Expectancy table showing percentages of students in each concept mapping post-test score range (scored using the new scoring method) who received various multiple-choice post-test scores.

Concept mapping post-test score ranges	Multiple-choice post-test scores								
	0	1	2	3	4	5	6	7	8
0-2	0	0	0	0	0	0	0	0	0
3-5	0	0	0	0	5.5	0	16.7	16.7	0
6-8	0	0	0	0	5.5	27.7	5.5	5.5	16.7
9-11	0	0	0	0	5.5	5.5	0	0	5.5
12-14	0	0	0	0	0	5.5	0	0	0

Table 18. Expectancy table showing percentages of students in each concept mapping post-test score range (scored using the modified Novak and Gowin scoring method) who received various multiple-choice post-test (plus unit test questions) scores.

Concept mapping post-test score ranges	Multiple-choice post-test plus unit test questions scores											
	0	1	2	3	4	5	6	7	8	9	10	11
16-20	0	0	0	0	0	0	0	0	0	5.5	0	0
21-25	0	0	0	0	0	0	5.5	0	0	5.5	0	0
26-30	0	0	0	0	0	0	0	11.1	5.5	0	5.5	0
31-35	0	0	0	0	0	0	0	5.5	16.7	5.5	0	0
36-40	0	0	0	0	0	0	5.5	0	0	0	5.5	0
41-45	0	0	0	0	0	0	0	0	5.5	0	0	0
46-50	0	0	0	0	0	0	5.5	0	0	0	5.5	0

Table 19. Expectancy table showing percentages of students in each concept mapping post-test score range (scored using the new scoring method) who received various multiple-choice post-test (plus unit test questions) scores.

Concept mapping post-test score ranges	Multiple-choice post-test plus unit test questions scores											
	0	1	2	3	4	5	6	7	8	9	10	11
0-2	0	0	0	0	0	0	0	0	0	0	0	0
3-5	0	0	0	0	0	0	5.5	5.5	5.5	11.1	0	0
6-8	0	0	0	0	0	0	5.5	5.5	22.2	5.5	16.7	0
9-11	0	0	0	0	0	0	5.5	5.5	0	0	5.5	0
12-14	0	0	0	0	0	0	0	0	0	0	0	0

Table 20. Expectancy table showing percentages of students in each concept mapping pretest score range (scored using the modified Novak and Gowin scoring method) who received various concept mapping pretest rank ranges.

Concept mapping pretest score ranges	Concept mapping pretest rank ranges				
	0-3	4-7	8-11	12-15	16-19
16-20	0	0	0	5.5	0
21-25	0	5.5	0	5.5	5.5
26-30	0	5.5	5.5	5.5	11.1
31-35	5.5	5.5	16.7	0	5.5
36-40	0	5.5	0	5.5	0
41-45	5.5	0	0	0	0

Table 21. Expectancy table showing percentages of students in each concept mapping pretest score range (scored using the new scoring method) who received various concept mapping pretest rank ranges.

Concept mapping pretest score ranges	Concept mapping pretest rank ranges				
	0-3	4-7	8-11	12-15	16-19
0-2	0	5.5	0	5.5	5.5
3-5	0	0	11.1	11.1	16.7
6-8	11.1	16.7	11.1	5.5	0
9-11	0	0	0	0	0
12-14	0	0	0	0	0

Table 22. Expectancy table showing percentages of students in each concept mapping post-test score range (scored using the modified Novak and Gowin scoring method) who received various concept mapping post-test rank ranges.

Concept mapping post-test score ranges	Concept mapping post-test rank ranges				
	0-3	4-7	8-11	12-15	16-19
16-20	0	0	0	0	5.5
21-25	0	0	5.5	0	5.5
26-30	0	0	11.1	5.5	5.5
31-35	5.5	5.5	11.1	5.5	0
36-40	0	5.5	0	11.1	0
41-45	5.5	0	0	0	0
46-50	5.5	5.5	0	0	0

Table 23. Expectancy table showing percentages of students in each concept mapping post-test score range (scored using the new scoring method) who received various concept mapping post-test rank ranges.

Concept mapping post-test score ranges	Concept mapping post-test rank ranges				
	0-3	4-7	8-11	12-15	16-19
0-2	0	0	0	0	0
3-5	0	5.5	5.5	5.5	11.1
6-8	16.7	11.1	11.1	11.1	5.5
9-11	0	5.5	5.5	5.5	0
12-14	0	0	0	0	0

Generally, the low validity coefficients may be explained by the reliability of the concept mapping and the multiple-choice pre- and post-tests. The reliability of the criterion measure influences the validity coefficient, therefore, anything affecting the reliability of the criterion measure also affects the validity coefficient (Sax, 1989). In fact, "the validity of a particular test can never exceed the reliability of the test" (Salvia & Ysseldyke, 1991, p. 154) The low validity coefficients from Table 13 were not unexpected considering the low reliability coefficients reported in the internal consistency section of this chapter.

In addition, the magnitude of a correlation is affected by the range of the individual student differences. When there is a restriction in this range the result can be an attenuation of the validity coefficient (Brown, 1983; Sax, 1989). In the case of the pretests, practically everyone scored under 50%, which restricted the range of scores. This was most likely because students had not yet received instruction on the material covered by the pretests. Likewise, practically everyone scored over 50% on the post-tests, which also restricted the range of scores. All students received the same instruction, which made them a homogeneous group, and probably caused the scores to cluster together. This restriction of the range of the criterion measure (the multiple-choice pre - and post-test) scores, very likely had a negative influence on the validity coefficients.

A problem already identified with the modified Novak and Gowin scoring method may explain the negative correlations between the concept mapping ranks and the concept mapping scores. As discussed before, a map showing a linear organization of concepts could score higher than a map showing many branches because hierarchical levels are awarded 5 points each. Although a linear organization may be appropriate for some content area, this was not the case for the target concept chosen for Study II. Therefore, although concept maps with linear organizations may have scored high using the modified Novak and Gowin scoring method, the expert may have ranked these same maps low because the concepts were more accurately organized with a branching structure. In this case, a negative correlation would be expected.

With regard to the concept mapping ranks assigned by an expert, biases may have also affected the validity coefficients (Brown, 1983). The expert assigning the ranks to the concept maps may have based the ranking on things other than knowledge of measurement from an ISE perspective. For example, many of the concept maps were disorderly and difficult to read. In these cases, the expert may have unintentionally allowed the poor structure of the concept maps affect the assigned ranks.

In addition, when validity coefficients are established for a small sample size, sampling errors are large, resulting in an attenuation of the validity coefficient (Crocker & Algina, 1986). The sample size for the concurrent validity study (n=18) likely contributed to the low validity coefficient. When the sample size was considered in conjunction with the low reliability of the multiple-choice tests, the restriction of range of the pre-and post-tests, and the biases that may have affected the concept map rankings, the low validity coefficients was not surprising.

Construct validity

Construct validity is established by examining hypothesized relationships between subject performance on the instrument and predicted performance on other criterion measures. Novak and Gowin (1983) argued for the construct validity of concept mapping based on the appearance of maps resulting from the concept mapping procedure and Ausubel's learning theory. Their argument is as follows.

> To the extent that Ausubelian learning theory validly describes cognitive learning, and to the extent that our concept mapping procedures are consistent with these learning principles for achievement assessment, we believe the bias in our procedures is not deleterious. Concept maps may be said to have construct validity in terms of evaluation theory. There is a correspondence between assessment of cognitive performance and what our theory predicts should be the cognitive organization resulting from meaningful learning. (p.105)

Novak and Gowin (1983) make the argument that the map resulting from the concept mapping procedure has the same characteristics as one would expect a cognitive structure to have given Ausubel's learning theory. Ausubel (1963) describes an existing cognitive structure as a stable body of general concepts and principles and is the major factor affecting meaningful learning, retention, and problem solving. Furthermore, Ausubel (1963) describes this structure as being hierarchical, with the most general and inclusive ideas occupying a position at the top of the structure and becoming progressively less inclusive toward the bottom. Mappers construct concept maps by following the concept mapping procedure and produce a structure that is organized from the most general concepts at the top of the hierarchy to the most specific concepts at the bottom, resembling the conceptual cognitive structure described by Ausubel.

Ausubel (1963) also asserts that as learning occurs, this structure changes to accommodate new information. Therefore, it was hypothesized that if learning occurred as a result of instruction between the concept mapping pre-and post-tests, then there would be a significant difference between the pre- and post-test scores. The multiple-choice test was also given before and after instruction to confirm that learning had occurred. Because the concept-mapping and multiple-choice tests were given at the same time, a repeated measures design was used to determine if test order had any effect on the pre- and post-test scores. The results of the repeated measures design for the concept mapping test scores derived from the modified Novak and Gowin scoring method, the concept mapping test scores derived from the new scoring method, and the multiple-choice test scores appear in Tables 24-26 respectively.

Table 24. Repeated measures analysis of the concept mapping tests (scored using the modified Novak and Gowin scoring method) by test order and time of the test.

	df	SS	MS	F
Order	1	11.78	11.78	.15
Error (between)	15	1187.46	79.16	
Time	1	24.82	24.82	.66
Time x Order	1	5.40	5.40	.14
Error (within)	15	563.12	37.54	

Table 25. Repeated measures analysis of the concept mapping tests (scored using the new scoring method) by test order and time of the test.

	df	SS	MS	F
Order	1	1.08	1.08	.19
Error (between)	15	85.86	5.72	
Time	1	27.53	27.53	8.34*
Time x Order	1	.36	.36	.11
Error (within)	15	49.52	3.30	

* p = .011

Table 26. Repeated measures analysis of the multiple-choice tests by test order and time of the test.

	df	SS	MS	F
Order	1	.41	.41	.51
Error (between)	15	12.06	.80	
Time	1	61.08	61.08	27.06*
Time x Order	1	.14	.14	.06
Error (within)	15	33.86	2.26	

* p = .000

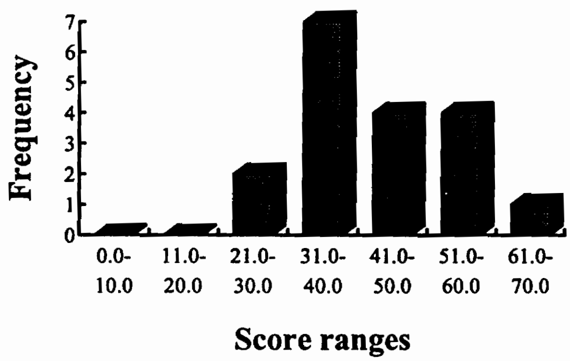
A significant difference in the multiple-choice pre- and post-test scores confirmed learning occurred among the students between the time the pre- and post-tests were given (see Table 26). There was no significant difference in the concept mapping pre- and post-test scores when the modified Novak and Gowin scoring method was used, suggesting the concept mapping test was not sensitive to changes in students as a result of instruction (see Table 24). Yet, when scored with the new scoring method, there was a significant difference between concept mapping pre-and post-test scores (see Table 25). These findings support the need for a more reliable scoring method for concept mapping. The modified Novak and Gowin method had problems as described earlier. The new scoring method, although somewhat more reliable (as shown in the validity coefficient of the concept mapping pretest), may also need modifications.

The repeated measures analysis showed the order in which the concept mapping and multiple-choice tests were taken had no significant effect on the test scores. In addition, there was no significant interaction between the order in which the concept mapping and multiple-choice tests were taken and the time they were taken.

Another way to examine the construct validity of a measure is to review the evidence from a criterion-related validity study (Crocker & Algina, 1986). Assuming the multiple-choice tests were a measure of knowledge of measurement from an ISE perspective, then if the concept mapping tests measured the same construct, they should have correlated positively with the multiple-choice tests. As was discussed in the criterion-related validity section above, there were very weak correlations between the concept mapping and multiple-choice tests.

Mapping Ability

The practice concept maps of the University Library were used to determine the quality of students' concept mapping skills. The topic chosen for this practice map was considered to be one all students would know well. Therefore, differences in concept map scores would likely be due to concept mapping ability rather than understanding of the target concept. There was a wide range in scores (range = 21 to 61, $n = 18$, and mean = 42.11) and the standard deviation was rather large (S.D. = 10.78), suggesting there was a wide range of concept mapping ability (see Figure 9). The most common error the scorers saw was the students' inability to hierarchically order the related concept words. For example, one map showed "copy card" as a more general topic than "copier." Another common problem occurred when the students tried to construct efficient concept maps. For example, one student connected several concepts to one point and instead of labeling that point with a concept word, the student simply wrote a description of how all the concepts were related (the "wagon spoke" effect). This was a very efficient way to show the interrelationships the student understood to exist between concepts, but this approach varied so much from the instructions, the map's score was lowered. Students' inability to make accurate concept maps influenced the data generated in this study. Although students appeared to quickly understand the basics of concept mapping, they did not develop their concept mapping skills to the point that they had sufficient concept mapping skills to produce maps that accurately reflected their knowledge.



Mean	42.11
Standard Deviation	10.78
Standard Error Mean	2.54
Minimum	21.00
Maximum	61.00
N	18.00

Figure 9. Distribution of concept mapping practice exercise scores for Study II.

CHAPTER 4

Conclusions and Discussion

The purpose of this study was to examine the reliability and validity of concept mapping as a pretest to assess prior knowledge for instructional decision making. If concept mapping is a valid measure of prior knowledge, it could provide data instructors could use to understand student entry knowledge and facilitate the instructional design process. This study was investigated the validity of concept mapping.

A well-documented reliability and validity study describes the reliability and validity evidence of a measure and allows the user to judge the measure's adequacy (Moss, 1994). Because a measure's validity depends on its purpose, "One can never reach the simple conclusion that a particular test is valid" (Cronbach, 1971, p. 447). The results of this study provide evidence that can be used to determine if concept mapping is reliable and valid for judging prior knowledge. However, the reliability and validity evidence gathered for this study strongly suggest that concept mapping and the concept mapping scoring method (as described by Novak & Gowin, 1993) are not valid measures of students' prior knowledge and should not be used as a data source for instructional decision making.

Conclusions

On the basis of the data collected to determine the reliability and validity of concept mapping, the concept mapping procedure (as described by Novak & Gowin, 1993) is unreliable and invalid. These conclusions are elaborated below.

Reliability

The reliability analyses included interrator reliability and internal consistency. Both analyses demonstrated concept mapping was an unreliable measure. The interrator reliability indicated that the concept mapping tests could not be scored reliably. The subjectivity required of the scorers by the available scoring methods contributed to the difficulty in achieving an acceptable interrator reliability. The internal consistency analysis indicated that items on the concept mapping test were inconsistent, meaning the test items were not measuring the same construct. The multiple-choice test was also unreliable, probably due to the poor quality of the items. Although the interrator reliability was very high for the multiple-choice test, the internal consistency was low. The increase in the reliability coefficient of the multiple-choice post-test when additional questions were added demonstrated that this was very likely the case. Overall, the results of the reliability analyses showed that concept mapping and the multiple-choice test used in this study were unreliable.

Validity

Four analyses were conducted to determine the validity of concept mapping. The content validity analysis showed very positive results. Experts rated each item very high in four areas relating to content validity, and the qualitative data confirmed the experts' high ratings. The face validity analysis showed that although about half of the students preferred concept mapping tests over multiple-choice tests as a measure of what they know about measurement from an ISE perspective, they had little confidence in the ability of either measurement device to assess their knowledge. The criterion-related validity (concurrent validity) showed no significant relationship between the scores on the concept mapping and multiple-choice tests, regardless of the scoring method used on the concept mapping tests.

The construct validity analysis revealed no significant difference in the order the concept mapping and multiple-choice tests were taken, and no significant interaction between the order and the time (pre- or post-test). However, there was a significant difference between the multiple-choice pre- and post-tests and the concept mapping pre- and post-tests when the concept mapping tests were scored with the new scoring method (there was no significant difference between the pre- and post-test scores when the concept mapping tests were scored with the Novak and Gowin method). The significant increase between pre-and post-tests for both measures should indicate that measure was sensitive to changes that occurred among the students between the pre- and post-test; but, because both measures were unreliable, it is unclear what the data mean.

The results of the validity analyses clearly indicated poor validity for both the concept mapping and multiple-choice tests. This was not unexpected given the poor reliability of both measures.

Discussion

Traditional methods of estimating reliability and validity of measures were difficult to apply to concept mapping, a nontraditional measurement device examined in this study. For example, a new scoring method had to be created to analyze the internal consistency of the concept mapping test. The scoring method recommended by Novak and Gowin was not usable for an internal consistency analysis because specific items were not identified for scoring according to this procedure. The new scoring procedure identified items that could be scored on the concept mapping tests. Perhaps alternative approaches for analyzing reliability and validity of non-traditional measurement devices would provide different results than those of this study.

Moss (1994) discussed the problems associated with depending on traditional psychometric approaches to establishing reliability and validity of a test: "Current conceptions of reliability and validity in educational measurement constrain the kinds of assessment practices that are likely to find favor (in a search for measurement devices to accomplish desired educational change), and these in turn constrain educational opportunities for teachers and students" (Moss, 1994, p. 10). Moss argued that traditional processes for establishing interrator reliability are not appropriate for many non-traditional measures of performance which are gaining attention in educational circles (e.g. portfolio assessments). Moss discussed a hermeneutic approach to assessing performance where several scorers assess multiple performance indicators (preferably selected by the subject) and then meet to discuss their interpretations. This process results in a group decision regarding the performance of the subject; generalizations regarding performance are therefore made by several judgments of multiple performances.

Perhaps the hermeneutic approach to analyzing concept maps as measures of prior knowledge as suggested by Moss (1994) may result in a more reliable assessment of concept mapping results. It could be that the traditional psychometric processes used in this research to evaluate the reliability of concept mapping are not appropriate for this kind of non-traditional measure/device. Although scoring procedures were developed to make traditional procedures possible, it is unclear what results would be found if different, proven, and acceptable approaches to reliability and validity were available.

Future studies which may build upon this study will have to overcome several problems. First, a reliable scoring method was not available for scoring concept maps.

Although progress was made in this study toward increasing the reliability of scoring, continued study and modifications are necessary. By improving reliability through an improved scoring method, more satisfactory validity may be possible because reliability defines the upward bound of validity (Sax, 1989).

Another problem experienced in this study was students' poor mapping abilities. More practice with feedback was needed for subjects to attain mastery of concept mapping. Research into concept mapping training and follow-up practice requirements are recommended before future reliability and validity studies are pursued. In a similar vein, the variables affecting students' mapping ability should also be examined. These variables may include cognitive style, creativity, and test-taking skills. Level of knowledge may also be a variable to consider; subjects in Study I and Study II said they felt they could not do the concept mapping exercise because they did not know enough about the topic. Investigations of these, and other, variables affecting concept mapping ability may give additional insight on the most effective concept mapping training procedure. In addition, just as test-taking skills can be improved for any test format (Sherman, 1984), it is likely concept mapping skills may be developed for improved performance and research in these skills should be pursued.

Finding subjects was also difficult. Because of the need for instructors to donate classroom time so students could be trained to do concept mapping and then complete the testing materials, willing subjects were difficult to locate. But, perhaps with further research to understand the most effective and efficient ways to learn to concept map, a mapping procedure that all subjects can use with minimal instruction and interruption to the normal class schedule can be developed.

The problems experienced during this study and their effects on the results should not prevent future researchers from looking for ways to increase the reliability and validity of concept mapping. As stated by the students and experts in Study II, concept mapping seemed to provide different information from a multiple-choice test. Students liked the concept mapping test because they could explain their understanding of concepts. Students also said explanation was not possible with multiple-choice tests, especially when their preferred answer was not among the item choices. Experts made similar comments in that it was clear from the concept maps why subjects' connected concepts incorrectly because students described the relationships between concepts. Once a reliable and valid concept mapping procedure is developed, future research should focus on the kinds of information gained from concept mapping compared to other kinds of instruments, and how this information can be used to benefit students during instructional planning.

The analysis of the construct validity data also suggests that concept mapping is worthy of further development and research. The concept maps showed a significant increase in scores between the pre- and post-test when the maps were scored by the new scoring method. This implies something happened between the pre- and post-test to significantly increase the scores. It is likely instruction and learning was that something; it may be that concept mapping is sensitive to changes in the subject as a result of instruction. For example, Beyerbach and Smith (1992) concluded students' cognitive structures become more like that of the instructor during instruction. Results in Study II are consistent with Beyerbach and Smith's conclusion as the items making up the key for the new scoring method were based on the expert map (constructed by the instructor). Therefore, the more the students' maps resembled the instructor's map, the higher the students' maps scored. A significant increase in map scores from the pre- to post-test indicates that the students' cognitive structures took on more of the characteristics of the instructor's cognitive structure. Further research into factors responsible for the significant increase between pre- and post- concept mapping scores is needed to understand this phenomenon.

Several reasons for choosing concept mapping as an alternative measurement device for measuring prior knowledge were given in Chapter 1. Some of these reasons were supported and argue for continuing research on concept mapping and others were not supported. One reason given for choosing concept mapping was that developing and implementing the procedure is straightforward (Novak & Gowin, 1993). The experts agreed that developing the concept mapping test was faster and less difficult compared to the multiple-choice test. And, the experts expressed a desire to continue learning about concept mapping so that they could use it in their own classrooms. Implementing concept mapping, however, was more difficult than anticipated. As mentioned before, finding willing subjects was difficult given the time required to participate in this study. Scoring the concept maps was also more difficult than expected because of the poor reliability of the scoring procedure recommended by Novak and Gowin.

Another proposed advantage of concept mapping was that students could represent their understanding of relationships between concepts freely. This was supported by comments made by students and experts. Students and experts agreed they liked the freedom students had to express understanding of concepts on the concept mapping test. Experts liked the fact that they were able to see how students described relationships between concepts. They felt this gave them different information than could be gained by a multiple-choice test. On the down side, students expressed concerns about the subjectivity that could influence the concept map scorer.

The ease of data analysis of concept mapping was another reason for choosing it for this study. Novak and Gowin (1993) and Jonassen et al. (1993) suggest that the data collected from a concept mapping procedure can be readily analyzed without complicated statistical methods and that a simple review of the maps provides a great deal of information. When asked if the scoring procedure was necessary, the experts replied that although scores would be nice for ranking students, scoring was not necessary to gain information for instructional planning. From a review of the concept mapping test results, the experts identified several areas in which students showed weaknesses that should be targeted during instruction. However, given the results of this study, the reliability and validity analyses should be done each time the concept mapping test is used until it is certain the concept mapping test is reliable and valid.

Just as the conclusion of this study cannot be that *all* multiple-choice tests are invalid and unreliable because the multiple-choice test used in this study was invalid and unreliable, the conclusion also cannot be that *all* concept mapping tests are invalid and unreliable. Although concept mapping in this study was invalid and unreliable, research to improve the reliability and validity of this measurement device and other Concept Mapping procedures should continue.

References

Anderson, R. C. (1977). The notion of schemata and the educational enterprise: General discussion of the conference. In R. Anderson, R. Spiro, & W. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, New Jersey: Lawrence Earlbaum Associates.

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York, New York: Grune & Stratton.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York, New York: Holt, Rinehart, and Winston.

Beyerbach, B. A., & Smith, J. M. (1992). Developing a profile of preservice teachers' thinking using concept mapping, stimulated recall, discourse analysis. *Journal of Research in Education*, 2(1), 60-67.

Brown, F. G. (1983). *Principles of educational and psychological testing*. New York, New York: Holt, Rinehart, and Winston.

Camp, R. R., Blanchard, P. N., & Huszczo, G. E. (1986). *Toward a more organizationally effective training strategy and practice*. Englewood Cliffs, New Jersey: Prentice Hall.

Coleman, S. L. & Sherman, T. M. (1992a, March). *Concept mapping: A tool for knowledge assessment.* Paper presented at the annual meeting of the Eastern Educational Research Association, Hilton Head, South Carolina.

Coleman, S. L. & Sherman, T. M. (1992b, April). *Concept mapping: A tool for knowledge assessment.* Poster presentation at the Virginia Tech Graduate Student Assembly Graduate Research Symposium, Blacksburg, Virginia.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* Fort Worth, Texas: Harcourt Brace Jovanovich.

Cronback, L. J. (1971). Test validation. In Robert L. Thorndike (Ed.), *Educational Measurement* (2nd ed.) (pp. 443-507). Washington, D.C.: American Council on Education.

Dick, W. & Carey, L. (1985). *The systematic design of instruction, 2nd ed.* Glenview, Illinois: Scott, Foresman & Co.

Donald, J. G. (1987). Learning schemata: Methods of representing cognitive, content and curriculum structures in higher education. *Instructional Science, 16*, 187-211.

Ferguson, G. A. & Takane, Y. (1989). *Stastical analysis in psychology and education.* (6th ed.). New York, New York: McGraw-Hill.

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1988). *Principles of instructional design.* Fort Worth, Texas: Holt, Rinehart, and Winston, Inc.

Gussarsky, E. & Gorodetsky, M. (1988). On the chemical equilibrium concept: Constrained word associations and conception. *Journal of Research in Science Teaching, 25*(5), 319-333.

Johnson, S. D. & Thomas, R. (1992). Technology education and the cognitive revolution. *Technology Teacher, 51*(4), 7-12.

Jonassen, D. H. (1987). Assessing cognitive structure: Verifying a method using pattern notes. *Journal of Research and Development in Education, 20*(3), 1-14.

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

Kerlinger, F. (1973). *Foundations of Behavior Research, 2nd ed.* New York, New York: Holt, Rinehart and Winston.

Lewis, L. (1991, April). *The development of understanding in elementary thermodynamics.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

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

McDonald, J. L. (1989). Accuracy and stability of cognitive structures and retention of geometric content. *Educational Studies in Mathematics, 20,* 425-448.

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology.* Fort Worth, Texas: Holt, Rinehart, and Winston.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision.* New York, New York: McGraw-Hill.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher, 23*(2), 5-12.

Novak, J. D. & Gowin, D. B. (1993). *Learning how to learn.* New York, New York: Cambridge University Press.

Novak, J. D. & Ridley, D. R. (1988). *Assessing student learning in light of how students learn.* Washington, D.C.: American Association for Higher Education. (ERIC Document Reproduction Service No. ED 299 923).

Okebukola, P. A., & Jegede, O. J. (1988). Cognitive preference and learning mode as determinants of meaningful learning through concept mapping. *Science Education, 72*(4), 489-500.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Popham, W. J. (1981). *Modern Educational Measurement.* Englewood Cliffs, New Jersey: Prentice-Hall.

Reiss, C., & Morley, R. (1993, April). *The harsh realities of ISD: Too many boxes, too little time or money.* Paper presented at the meeting of the National Society for Performance and Instruction, Chicago, Illinois.

Richey, R. (1986). *Theoretical and conceptual bases of instructional design.* New York, New York: Nicols Publishing.

Salvia, J., & Ysseldyke, J. E. (1991). *Assessment* (5th ed.). Boston, Massachusetts: Houghton Mifflin.

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation.* Belmont, California: Wadsworth Publishing.

Seels, B., & Glasgow, Z. (1990). *Exercises in instructional design.* Columbus, Ohio: Merrill Publishing.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology, 63*(3), 225-234.

Shavelson, R. J. (1974). Methods for examining representations of a subject-matter structure in a student's memory. *Journal of Research in Science Teaching, 11*(3), 231-249.

Tuckman, B. W. (1975). *Measuring educational outcomes: Fundamentals of testing.* New York, New York: Harcourt Brace Jovanovich, Inc.

APPENDIX 1

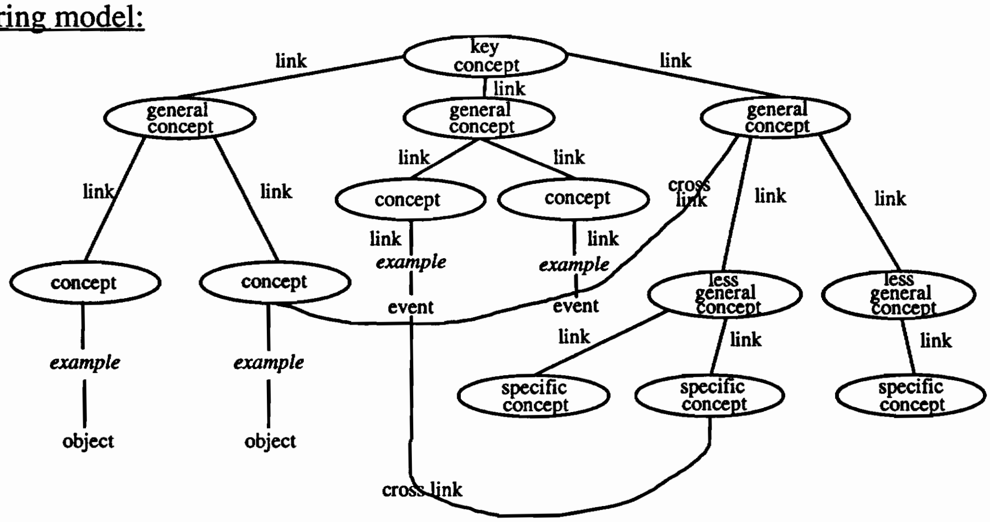
Instructions for scoring a concept map

Scoring procedure for concept maps
(Novak & Gowin, 1993, p. 36).

Concept maps are scored on three primary criteria. First, are correct relationships indicated between two concepts? Second, is there a hierarchy appearing in the concept map? And third, are there crosslinks among concepts in the concept map? Points are also awarded if the subject includes examples.

1. Propositions. Is the relationship between two concepts indicated by a connecting line and linking word(s)? Is the relationship correct? For each meaningful, correct proposition shown, score 1 point.
2. Hierarchy. Does the map show hierarchy? Is each subordinate concept more specific and less general than the concept drawn above it (in context of the material being mapped)? Score 5 points for each valid level of the hierarchy.
3. Crosslinks. Does the map show meaningful connections between one segment of the concept hierarchy and another segment? Is the relationship shown significant and correct? Score 10 points for each crosslink that is both correct and significant and 2 points for each crosslink that is correct but does not illustrate a synthesis between sets of related concepts or propositions.
4. Examples. Specific events or objects that are valid instances of those designated by the concept label are scored 1 point each.

Scoring model:



Scoring for this model:

Relationships (if valid)	= 14
Hierarchy (if valid) 4 x 5	= 20
Crosslinks (if valid and significant) 2 x 10	= 20
Examples	= 4
	58 points total

Modified scoring procedure for concept maps

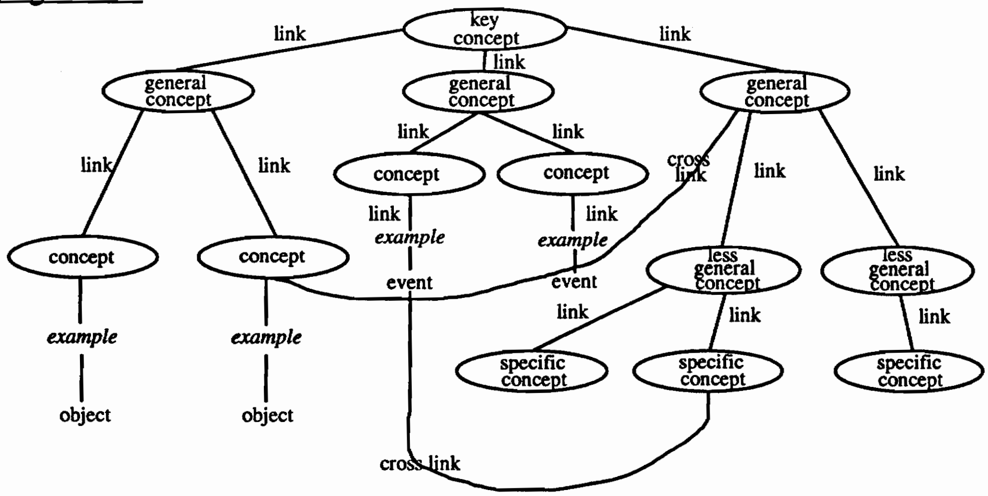
Concept maps are scored on three primary criteria. First, are correct relationships indicated between two concepts? Second, is there a hierarchy appearing in the concept map? And third, are there crosslinks among concepts in the concept map? Points are also awarded if the subject includes examples.

1. Relationships. Is the relationship between two concepts indicated by a connecting line and linking word(s)? Is the relationship correct? For each meaningful, correct proposition shown, score 1 point. If the relationships indicates an advanced, complex understanding of the relationship, score 2 points. If no relationship is stated, score no points.
2. Hierarchy. Does the map show hierarchy? Is each subordinate concept more specific and less general than the concept drawn above it (in context of the material being mapped)? Score 5 points for each valid level of the hierarchy. If within a branch no relationship is stated between one of the segments, do not count this level of the hierarchy. If within a branch, an incorrect relationship is stated between one of the segments, do not count this level of the hierarchy. On a map with multiple hierarchical branches, choose the longest branch and count the levels in that branch for the hierarchical score.
3. Cross links. Does the map show meaningful connections between one branch of the concept hierarchy and another branch? Is the relationship shown significant and correct? Score 2 points for each crosslink that is correct and score 4 points for each cross link that is both correct and illustrates an advanced, complex synthesis between sets of related concepts.
4. Examples. Specific events or objects that are valid instances of those designated by the concept label are scored 1 point each.
5. If the segments between concepts contain arrows, ignore the direction of the arrow and score as indicated above.
6. If the map shows the concepts in two different contexts (e.g., study as a verb in one relationship and study as a noun in another relationship) do not count off as long as the stated relationship or crosslink is correct.
7. If the map shows one branch with concepts arranged in a diamond pattern and the rest of the branch continuing under the bottom point of the diamond, select one side of the diamond to score for the hierarchical score. The two segments connecting to the concept making the fourth corner of the diamond should be considered crosslinks.
8. If the map contains a multi-spoked node (giving a "wagon wheel" appearance) without a relationship described on each branch , only those spokes that are labeled with a relationship are scored, and if the spoke is part of the branch and not labeled with a relationship, then it is not counted in the hierarchy. The node (where the segments converge) does not affect scoring. To count the levels of the hierarchy, the scorer will have to determine which path through the wagon wheel

makes the longest branch and count the levels in that branch. Any relationships connected to the branch should be considered crosslinks.

Scoring concept maps requires subjective judgment on the part of the scorer. The scorer must try to interpret what the mapper meant when unusual structures appear within a map. If the mappers intent was obviously not what was indicated by the structure, scorers should act on what they believe the mapper intended to convey.

<u>Scoring model:</u>



Scoring for this model:

Relationships (if valid)	= 14
Hierarchy (if valid) 4 x 5	= 20
Crosslinks (if valid and significant) 2 x 10	= 20
Examples	= 4
	58 points total

New scoring procedure for concept maps

The following is a list of paired concept words identified from an expert's concept map on measurement from an ISE perspective. Look for these pairs of concept words in each concept map. There is a match when the concept words on the map are directly linked (a crosslink is O.K.) — no concept word should appear between the paired concepts. Award one point for each pair appearing on a concept map demonstrating a correct relationship. If no relationship is described one point is still awarded. If the relationship described is incorrect, then no points are awarded.

Concept Mapping Test Items:

1. measurement from an ISE perspective - work measurement
2. measurement from an ISE perspective - organizational performance measures
3. organizational performance measures - organizational system
4. organizational system - processes
5. organizational system - inputs
6. organization system - outputs
7. work measurement - standard time
8. processes - tasks
9. inputs - efficiency
10. inputs - productivity
11. outputs - productivity
12. outputs - effectiveness
13. work measurement - efficiency
14. standard time - tasks

APPENDIX 2

Instructions for producing a concept map

Instructions for producing a concept map.

(Adapted from Jonassen et al., 1993 and Novak & Gowin, 1993).

1. Think about the key concept (the concept being mapped) and the list of concept words related to the key concept that have been given to you.
2. Rank the list in order of inclusiveness.
3. Write the key concept word at the top of the page and draw a circle around it.
4. Examine your list of concept words and select the most inclusive concept that is related to the key concept. Write this word below the key concept and draw a circle around it.
5. Consider the relationship between the two concepts on the page. Draw a line between these concept and write a brief (1-5 linking words) description to show how these concepts are related.
6. Select another concept from the list. Write this concept word on the page and circle it.
7. Identify any relationships between the first two concepts and the new concept on the concept map. Draw a line or lines between concepts that are related. Label each line to identify the type of relationship between the two concepts.
8. Continue this process until all the related concepts are included on the concept map.
9. Check the map to ensure all relationships between concepts are shown by labeled lines.
10. Check the organization of the concept map. Does it appear messy and confusing? If so, redraw the map to minimize confusion, but keep all lines showing relationships.

APPENDIX 3

Study I data collection materials

Consent Form

HOW INVOLVED	This project involves: 1. Being trained to do concept mapping. 2. Completing a multiple-choice pre-test on regression. 3. Completing a concept mapping pre-test on regression. 4. Completing a survey of your attitudes toward multiple choice and concept mapping as pre-tests of your knowledge of regression. This survey will also ask demographic questions. 5. Completing a multiple-choice post-test on regression. 6. Completing a concept mapping post-test on regression.
PRIVACY	The data collected for this project will be used to estimate the reliability and validity of concept mapping as a pre-test. To do so requires that your pre-testing materials, post-testing materials and demographic data be matched. You will be asked to write the last four digits of your social security number and the numbers indicating your parents' address (e.g., street number of the house) on all data-gathering documents. Your responses will not be attributed to you directly in any way. Also, the results of the pre- and post- tests will not affect your grade in this course.
BENEFITS	From this project we hope to be able to show that concept mapping is a valid and reliable pre-test.
WITHDRAW PROCESS	You are free to withdraw from this study at any time without penalty or prejudice, by contacting Tom Sherman, Division of Curriculum & Instruction, (231-5598), or Dr. Ernest R. Stout, Chair of the Institutional Review Board (231-9359).
CONTACT	This study has been approved by the Human Subjects Committee and the Instructional Review Board. If you have any further questions please contact Susan Coleman, Division of Curriculum & Instruction, (231-3191).

Your signature below indicates that you have read the information above and have agreed to participate in the research project.

______________________________ __________________

Signature of participant Date

Out-of-Class Practice Exercise

(Adapted from Jonassen et al., 1993 and Novak & Gowin, 1993).

Please do the following exercise at home and turn it in at the beginning of class next week. If you have any questions, please call me, Susan Coleman, 231-3191 during the day, or 961-3613 in the evening. Thank you again for your participation!

1. In the upper right hand corner, please write the last four digits of your social security number plus the numbers indicating your parents' address (e.g., house number). Your responses will remain confidential.

2. Think about the key concept, VPI library, and the list of concept words related to the key concept that have been given to you.

3. Rank the list in order of inclusiveness.

4. Write the key concept word at the top of the attached page and draw a circle around it.

5. Examine your list of concept words and select the most inclusive concept that is related to the key concept. Write this word below the key concept and draw a circle around it.

6. Consider the relationship between the two concepts on the page. Draw a line between these concepts and write a brief (1-5 linking words) description to show how these concepts are related.

7. Select another concept from the list. Write this concept word on the page and circle it.

8. Identify any relationships between the first two concepts and the new concept on the concept map. Draw a line or lines between concepts that are related. Label each line to identify the type of relationship between the two concepts.

9. Continue this process until all the related concepts are included on the concept map.

10. Check the map to ensure all relationships between concepts are shown by labeled lines.

11. Check the organization of the concept map. Does it appear messy and confusing? If so, redraw the map to minimize confusion, but keep all lines showing relationships.

Key concept: VPI Library

Related words:	Rank order:
copiers	1.
research	2.
librarian	3.
study	4.
journals	5.
circulation desk	6.
VTLS	7.
copy cards	8.
books	9.

Concept Mapping Out-of-Class Practice

In the upper right hand corner, please write the last four digits of your social security number and the numbers indicating your parents' address (e.g., house number).

Follow the directions on the cover page to produce a concept map of your understanding of the VPI library.

Concept Mapping Pre- and Post-test
(Adapted from Jonassen et al., 1993 and Novak & Gowin, 1993).

1. Think about the key concept, regression, and the list of concept words related to the key concept that have been given to you.

2. Rank the list in order of inclusiveness (from most general to most specific).

3. Write the key concept word at the top of the attached page and draw a circle around it.

4. Examine your list of concept words and select the most inclusive concept that is related to the key concept. Write this word below the key concept and draw a circle around it.

5. Consider the relationship between the two concepts on the page. Draw a line between these concepts and write a brief (1-5 linking words) description to show how these concepts are related.

6. Select another concept from the list. Write this concept word on the page and circle it.

7. Identify any relationships between the first two concepts and the new concept on the concept map. Draw a line or lines between concepts that are related. Label each line to identify the type of relationship between the two concepts.

8. Continue this process until all the related concepts are included on the concept map.

9. Check the map to ensure all relationships between concepts are shown by labeled lines.

10. Check the organization of the concept map. Does it appear messy and confusing? If so, redraw the map to minimize confusion, but keep all lines showing relationships.

Key concept: Regression

Related words:	Rank order:
coefficient	1.
t-test	2.
residual	3.
R^2	4.
intercept	5.
independent variable	6.
dependent variable	7.
least squares equation	8.
F test	9.
b_0	10.
b_1	11.
β_0	12.
$\rho^2 = 0$	13.
slope	14.
y	15.
$\hat{y}$	16.

Concept Mapping Pre- and Post-test

In the upper right hand corner, please write the last four digits of your social security number and the numbers indicating your parents' address (e.g., house number).

Follow the directions on the cover page to produce a concept map of your understanding of regression.

Multiple Choice Pre- and Post- Test

Circle the best answer for each item. You may use the attached formula sheet, if necessary.

1. Even if the Pearson Product Moment Correlation coefficient between two variables is zero, there may be a functional relationship between the two variables.

   1. true
   2. false

The following data represent the percentage of voluntary homework problems completed by each of 20 students and their final grade at the end of the course (converted to a 100-pt scale). Use this data to answer questions 2-6.

Problems: X	Grade: Y	Problem: X	Grade: Y
50	75	80	80
60	75	50	75
80	90	95	90
70	80	40	60
90	85	80	50
40	60	85	70
100	98	95	85
85	95	70	75
90	95	40	60
80	80	30	55

$\sum X = 1410$ $\sum Y = 1533$ $\sum XY = 112{,}350$

$\sum X^2 = 108{,}400$ $\sum Y^2 = 121{,}229$

$\overline{X} = 70.5$ $\overline{Y} = 76.65$

2. The correlation between amount of homework completed and final grade is:

   a. $r = -0.74$
   b. $r = -0.26$
   c. $r = 0.74$
   d. $r = 0.26$

3. Choose the statement that most accurately describes this correlation.
   a. The more homework someone completes, the lower the grade they are likely to receive
   b. About 54 percent of the variance in final grades is attributed to the amount of homework completed by students.
   c. There is a high, positive correlation between amount of homework completed and the final grade.
   d. 26% of the variance in final grades is attributed to the amount of homework completed by students.

4. Which test is most often used to determine if the correlation is significant?
   a. *F* test
   b. *t* test
   c. Tukey test
   d. Dunn's Multiple Comparison Test

5. The regression equation for predicting grade is:
   a. $\hat{Y} = 0.741X + 34.19$
   b. $\hat{Y} = 0.475X + 43.16$
   c. $\hat{Y} = 0.264X + 19.10$
   d. $\hat{Y} = 1.74X + 34.19$

6. The regression equation for predicting homework problems from grade is:
   a. $\hat{X} = 2.81Y + 1.07$
   b. $\hat{X} = 0.475Y + 43.16$
   c. $\hat{X} = 1.15Y + 17.65$
   d. $\hat{X} = 0.09Y + 6.091$

7. Within a group of 200 faculty members who have been at a well-known university for less than 15 years (that is, since before the salary curve levels off) the equation relating salary (in thousands of dollars) to years of service is $\hat{Y} = 0.9X + 15$. For 100 administrative staff at the same university, the equation is $\hat{Y} = 1.5X + 10$. Assuming that all differences are significant, interpret the meaning of these equations. How many years must pass before an administrator and a faculty member can earn roughly the same salary?
   a. 14 and 1/2 years
   b. 11 years
   c. 8 and 1/3 years
   d. 5 and 1/4 years

8. After working for the university for 12 years, what would you predict a faculty member's salary to be?
   a. \$41,600
   b. \$33,000
   c. \$25,800
   d. \$21,200

9. Consider the following regression equation:

   $$\hat{Y} = 1.813X_1 + .352X_2 + 1.069$$

   What is the intercept of this regression line?
   a. 1.813
   b. 1.069
   c. 2.557
   d. .744

10. How would you test the significance of the regression of Y on X_1 and X_2?
   a. F test
   b. t test
   c. r test
   d. χ^2 test

11. What is the null hypothesis when testing the significance of the regression of Y on X?
   a. $\overline{X} = \overline{Y}$
   b. $\overline{X} \neq \overline{Y}$
   c. The regression coefficient is not equal to zero.
   d. The regression coefficient is equal to zero.

12. A raw score regression coefficient of .50 is obtained when Y is regressed on X. This coefficient should be interpreted to mean that
   a. for each 10 point change on X, there is an associated 5.0 point change on Y.
   b. 50 percent of the variance in Y is explained by scores on X.
   c. 25 percent of the variance in Y is explained by scores on X.
   d. for each 10 point change on Y, there is associated a 5.0 point change on X.

13. When selecting variables for inclusion in a multiple regression equation, you should select variables that correlate highly with the criterion and:
   a. minimally among themselves.
   b. maximally among themselves.
   c. equally among themselves.
   d. not correlate at all among themselves.

14. As conventionally coded, which of the following variables would yield uninterpretable results if used in a multiple regression procedure to predict "marital satisfaction?"
   a. gender
   b. number of years married
   c. religious preference
   d. educational level

Formula Sheet

$$s_X = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

$$\text{cov}_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{N-1}$$

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}}$$

$$\hat{Y} = bX + a$$

$$r = \frac{SP_{XY}}{\sqrt{SS_X SS_Y}}$$

$$b = \frac{\text{cov}_{XY}}{s_Y^2} \text{ when } \hat{X} = bY + a$$

$$SS_X = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$SS_Y = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

$$SP_{XY} = \sum XY - \frac{(\sum X)(\sum Y)}{N}$$

$$b = \frac{\text{cov}_{XY}}{s_X^2} \text{ when } \hat{Y} = a + bX$$

$$Spearman's\ \rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$a = \overline{Y} - b\overline{X}$$

Student Questionnaire

In the upper right hand corner of this page, please write the last four digits of your social security number plus the numbers indicating your parents' address (e.g. house number). Please answer the following questions about the concept mapping and the multiple-choice testing procedures you completed. Your answers to these questions are important to understand how well each procedure assesses your understanding of a topic. Your answers will be kept anonymous, although your answers may be quoted and attributed to "a subject" in a report of this research. Thank you for your participation in this research project.

1. Which do you prefer for measuring what you know - concept mapping or multiple-choice tests? Why?

2. How was the information you provided on the concept map different from the information you provided on the multiple-choice test?

3. Do you feel confident that your current knowledge of regression was fairly assessed using the concept mapping procedure? Why or why not?

4. Do you feel confident that your knowledge of regression was fairly assessed using the multiple-choice test? Why or why not?

5. About how long did you work on your concept map of regression?

6. About how long did you work on your multiple-choice test of regression?

7. In general, what do you think about concept mapping as a pre-test?

<u>Demographic Information:</u>

1. Academic Level (circle one):
   master's student
   doctoral student
   non-degree student
2. Major: __________
3. Gender: __________
4. Age: _____

5. Have you ever used concept mapping before participating in this study?_____
   If yes, what did you map, and for what purpose?

6. How many semester credit hours of statistics classes have you <u>completed</u>? _____

7. a. In what year did you <u>complete</u> your most recent statistics class? _____
   b. Were you tested on regression in that statistics class? _____
   c. What was your course grade in that statistics class? _____

8. Have you *ever* been in a statistics class where you were tested on regression? _____

9. Have you ever used regression in a current or past job? _____
   If yes, please explain how you used regression in your job.

APPENDIX 4

Study II data collection materials

Subject Consent Form

HOW INVOLVED	This project involves: 1. Being trained to do concept mapping. 2. Completing a multiple-choice pre-test on measurement. 3. Completing a concept mapping pre-test on measurement. 4. Completing a survey of your attitudes toward multiple choice and concept mapping as pre-tests of your knowledge of measurement. This survey will also ask demographic questions. 5. Completing a multiple-choice post-test on measurement. 6. Completing a concept mapping post-test on measurement.
PRIVACY	The data collected for this project will be used to estimate the reliability and validity of concept mapping as a pre-test. To do so requires that your pre-testing materials, post-testing materials and demographic data be matched. You will be asked to write the last four digits of your social security number and the numbers indicating your parents' address (e.g., street number of the house) on all data-gathering documents. Your responses will not be attributed to you directly in any way. Also, the results of the pre- and post- tests will not affect your grade in this course.
BENEFITS	From this project we hope to be able to show that concept mapping is a valid and reliable pre-test.
WITHDRAW PROCESS	You are free to withdraw from this study at any time without penalty or prejudice, by contacting your instructor, Garry Coleman (231-7420), Tom Sherman, Division of Curriculum & Instruction (231-5598), or Dr. Ernest R. Stout, Chair of the Institutional Review Board (231-9359).
CONTACT	This study has been approved by the Human Subjects Committee and the Instructional Review Board. If you have any further questions please contact Susan Coleman, Division of Curriculum & Instruction, (231-3191).

Your signature below indicates that you have read the information above and have agreed to participate in the research project.

______________________________ ________________

Signature of participant Date

Expert Consent Form

HOW INVOLVED

This project investigates the reliability and validity of concept mapping as a pre-test. Your participation includes examining the concept mapping and multiple choice tests to estimate the content validity of each test, ranking subjects' maps as part of the estimation of the criterion-related validity, and being interviewed regarding your opinions of concept mapping as a pre-test.

PRIVACY

Because only two experts are involved in this study and the content of the course from which subjects come is also reported, there is a remote chance that you could be identified as part of this project. However, your responses will not be attributed to you directly. In the final report, you will not be identified by name, and will be referred to as "Expert A" or "Expert B".

BENEFITS

From this project we hope to be able to show that concept mapping is a valid and reliable pre-test. Your responses to the interview questions will provide important data for estimating the validity of concept mapping as a pre-test.

WITHDRAW PROCESS

You are free to withdraw from this study at any time without penalty or prejudice, by contacting Tom Sherman, Division of Curriculum & Instruction, (231-5598), or Dr. Ernest R. Stout, Chair of the Institutional Review Board (231-9359).

CONTACT

This study has been approved by the Human Subjects Committee and the Instructional Review Board. If you have any further questions please contact Susan Coleman, Division of Curriculum & Instruction, (231-3191).

Your signature below indicates that you have read the information above and have agreed to participate in the research project.

__________________________________ __________________

Signature of participant Date

Questionnaire for Content Validity

The purpose of content validity is to assess whether the items on the multiple choice and concept mapping tests adequately represent the domain of measurement from an ISE perspective, as presented in the Introduction to ISE course.

The domain of measurement from an ISE perspective is defined by the student performance objectives:
The student will be able to

1. Define work measurement, organizational performance measurement, and standard time.
2. Define effectiveness, efficiency, and productivity.
3. Explain the relationship between work measurement and organizational performance measurement.
4. Explain the relationship between effectiveness, efficiency, and productivity.

(Assume all objectives are weighted equally.)

A. Consider the following questions regarding the student performance objectives:

1. Please identify any objectives that go beyond the domain of measurement (from an ISE perspective, as presented in the Introduction to ISE course).

2. Please add any objectives that you think should be added to fully represent the domain of measurement (from an ISE perspective, as presented in the Introduction to ISE course). Use the back of this page if necessary.

B. In the matrix, indicate how well each test item matches the objectives in terms of characteristics shown in the top row. These characteristics have been defined below. Use the following scale when making your judgment:

1 = poor match 2 = fair match 3 = good match 4 = excellent match

subject matter - Does the content of the item match the content of the objective?
level of complexity of the performance required - Is the required complexity of performance as indicated by the objective similar to that required by the test item?
format of presentation - Does the question type require the respondent to display a behavior beyond that required by the objective (e.g., does the question use vocabulary at a level beyond what the respondent is expected to have)?
mode of response - Does the nature of the response require skills beyond that required by the objective (e.g., does an answering the question require a mathematical skill beyond those specified or implied in the objectives)?

multiple choice test item number:	subject matter	level of complexity of the performance required	format of presentation	mode of responses
1				
2				
3				
4				
5				
6				
7				
8				

concept mapping item	subject matter	level of complexity of the performance required	format of presentation	mode of responses
work measurement				
organizational performance measurement				
organizational system				
efficiency				
effectiveness				
productivity				
standard time				
inputs				
outputs				
processes				
tasks				

C. Overall, is there a balance between the emphasis on subcategories (i.e., those concepts related to measurement from an ISE perspective) within the objectives and the emphasis on subcategories as reflected by test items?

Practice Exercise

1. In the upper right hand corner, please write the last four digits of your social security number plus the numbers indicating your parents' address (e.g., house number). Your responses will remain confidential.

2. Think about the key concept, VPI library, and the list of concept words related to the key concept that have been given to you.

3. Rank the list in order of inclusiveness.

4. Write the key concept word at the top of the attached page and draw a circle around it.

5. Examine your list of concept words and select the most inclusive concept that is related to the key concept. Write this word below the key concept and draw a circle around it.

6. Consider the relationship between the two concepts on the page. Draw a line between these concepts and write a brief (1-5 linking words) description to show how these concepts are related.

7. Select another concept from the list. Write this concept word on the page and circle it.

8. Identify any relationships between the first two concepts and the new concept on the concept map. Draw a line or lines between concepts that are related. Label each line to identify the type of relationship between the two concepts.

9. Continue this process until all the related concepts are included on the concept map.

10. Check the map to ensure all relationships between concepts are shown by labeled lines.

11. Check the organization of the concept map. Does it appear messy and confusing? If so, redraw the map to minimize confusion, but keep all lines showing relationships.

Key concept: VPI Library

Related words:	Rank order:
copiers	1.
research	2.
librarian	3.
study	4.
journals	5.
circulation desk	6.
VTLS	7.
copy cards	8.
books	9.

Concept Mapping Practice

In the upper right hand corner, please write the last four digits of your social security number and the numbers indicating your parents' address (e.g., house number).

Follow the directions on the cover page to produce a concept map of your understanding of the VPI library.

Concept Mapping Pre- and Post-test
Measurement from the ISE perspective

1. Think about the key concept, measurement (from the perspective of an industrial engineer), and the list of concept words related to the key concept that has been given to you.

2. Rank the list in order of inclusiveness (from most general to most specific).

3. Write the key concept word at the top of the attached page and draw a circle around it.

4. Examine your list of concept words and select the most inclusive concept that is related to the key concept. Write this word below the key concept and draw a circle around it.

5. Consider the relationship between the two concepts on the page. Draw a line between these concepts and write a brief (1-5 linking words) description to show how these concepts are related.

6. Select another concept from the list. Write this concept word on the page and circle it.

7. Identify any relationships between the first two concepts and the new concept on the concept map. Draw a line or lines between concepts that are related. Label each line to identify the type of relationship between the two concepts.

8. Continue this process until all the related concepts are included on the concept map.

9. Check the map to ensure all relationships between concepts are shown by labeled lines.

10. Check the organization of the concept map. Does it appear messy and confusing? If so, redraw the map to minimize confusion, but keep all lines showing relationships.

Key concept: measurement (from the perspective of an industrial engineer)

Related words:	Rank order:
work measurement	1.
tasks	2.
processes	3.
organizational performance measurement	4.
outputs	5.
efficiency	6.
inputs	7.
standard time	8.
productivity	9.
organizational system	10.
effectiveness	11.

Concept Mapping Pre- and Post-test

In the upper right hand corner, please write the last four digits of your social security number and the numbers indicating your parents' address (e.g., house number).

Follow the directions on the cover page to produce a concept map of your understanding of measurement, (from the perspective of an industrial engineer).

Multiple Choice Pre- and Post-test
Measurement from the ISE perspective.

Please write the last four digits of your social security number and the numbers making up your parents address (e.g., house number) in the upper right hand corner. For each item, please circle the best response.

1. Work measurement is best described by which of the following statements.
   a. Using a time clock or time sheet to record hours worked.
   b. Establishing a standard time for performing a task.
   c. Analyzing motions to improve the efficiency of how a task is performed.
   d. Observing production times using a stop watch.
   e. Determining the productivity of a manufacturing plant.

2. Select the statement that best exemplifies organizational performance measurement.
   a. Monitoring the instrument panel of a nuclear power plant.
   b. Tracking the price of a company's stock.
   c. Calculating the production efficiency of a brewing process.
   d. Comparing organizational income to organizational expenses.
   e. Assigning grades to the students in a class.

3. Effectiveness is . . .
   a. Actual outputs compared to expected outputs.
   b. Resources expected to be consumed compared to resources actually consumed.
   c. The ratio of outputs over inputs.
   d. The ratio of revenues over expenses.
   e. The ratio of actual outputs compared to resources expected to be consumed.
   f. Widgets produced per hour compared to a standard.

4. Efficiency is . . .
   a. Actual outputs compared to expected outputs.
   b. Resources expected to be consumed compared to resources actually consumed.
   c. The ratio of outputs over inputs.
   d. The ratio of revenues over expenses.
   e. The ratio of actual outputs compared to resources expected to be consumed.
   f. Widgets produced per hour compared to a standard.

5. Which of the following IS NOT a measure of productivity?
   a. tons of coal per man-day
   b. automobiles per labor cost plus material cost
   c. barrels of beer per thousand cubic feet of natural gas
   d. sales revenue per square foot of floor space
   e. student credit hours taught per full-time faculty member

6. The difference between standard time and normal time is . . .
   a. Normal time is based on actual time to complete a task, standard time is not.
   b. Standard time is based on the standard pace, normal time is based on the normal pace.
   c. Normal time is the average from a normal curve of task times, standard time is based on production quotas.
   d. Standard time includes allowances for personal needs and delays, normal time does not.
   e. Normal time includes a performance rating, standard time does not.

7. The difference between work measurement and organizational performance measurement is most like the difference between . . .
   a. brain surgery and amputation.
   b. cholesterol level and overall health.
   c. miles per hour and miles per gallon.
   d. weight loss and getting into shape.
   e. outdoor air temperature and weather forecast.

8. Which of the following statements is most likely to be true?
   a. If an organization is effective and efficient, it is also productive.
   b. If an organization is productive and effective, it is also efficient.
   c. If an organization is productive and efficient, it is also effective.
   d. If an organization is profitable, it must be productive and efficient.
   e. If an organization is effective, it must be profitable.

Questions added to the post-test to increase internal consistency:

For each of the example measures listed below, identify each performance criterion being measured most directly: effectiveness, efficiency, productivity, quality checkpoint #1, quality checkpoint #2, quality checkpoint #3, quality checkpoint #4, or quality checkpoint #5.

9. $\frac{\text{Standard Hours}}{\text{Direct Hours Worked}}$ ________________________

10. $\frac{\text{Units Produced}}{\text{Units Scheduled}}$ ____________________

11. $\frac{\text{\# Software Instructions Written}}{\text{\# Software Engineers}}$ ____________________

Student Questionnaire

In the upper right hand corner of this page, please write the last four digits of your social security number plus the numbers indicating your parents' address (e.g. house number). Please answer the following questions about the concept mapping and the multiple-choice testing procedures you completed. Your answers to these questions are important to understand how well each procedure assesses your understanding of a topic. Your answers will be kept anonymous, although your answers may be quoted and attributed to "a subject" in a report of this research. Thank you for your participation in this research project.

1. Which do you prefer for measuring what you know - concept mapping or multiple-choice tests? Why?

2. How was the information you provided on the concept map different from the information you provided on the multiple-choice test?

3. Do you feel confident that your current knowledge of measurement (from an industrial engineering perspective) was fairly assessed using the concept mapping procedure? Why or why not?

4. Do you feel confident that your knowledge of measurement (from an industrial engineering perspective) was fairly assessed using the multiple-choice test? Why or why not?

5. About how long did you work on your concept map of measurement?

6. About how long did you work on your multiple-choice test of measurement?

7. In general, what do you think about concept mapping as a way for an instructor to assess what you know prior to instruction with the intention of using this information to plan instruction (NOT to assign grades to students)?

Demographic Information:

1. Academic Level (circle one):
   - freshman
   - sophomore
   - junior
   - senior
   - graduate student

2. Major: __________
3. Gender: __________
4. Age: _____

5. Have you ever used concept mapping before participating in this study? _____
   If yes, what did you map, and for what purpose?

6. How many semester credit hours of industrial engineering classes have you completed? _____

7. a. In what year did you complete your most recent industrial engineering class? _____
   b. Were you tested on measurement in that class? _____
   c. What was your course grade in that class? _____

8. Have you *ever* been in an engineering class where you were tested on measurement?
   ____

9. Have you ever used measurement (from an industrial engineer's perspective) in a current or past job? __________ If yes, please explain how.

Expert Interview Questions

1. Did the concept mapping procedure measure the concepts it was supposed to measure? Why or why not? How or how not?

2. Which do you prefer as a measure of what your students know (for the purpose of instructional planning) - concept maps or multiple-choice tests? Why?

3. In what ways was the information your students provided on the concept map different from the information they provided on the multiple-choice test?

4. Do you feel confident that your students' knowledge of measurement (from an ISE perspective) was fairly assessed using the concept mapping procedure? Why or why not?

5. Do you feel confident that your students' knowledge of measurement (from an ISE perspective) was fairly assessed using the multiple choice test? Why or why not?

6. Would you use the concept mapping procedure as tool for instructional planning?

7. Were you able to identify anything from the concept mapping pretests that would suggest areas to target during instruction?

8. Do you think a scoring procedure is necessary in order for you to make decisions for instructional planning?

9. Which was easier to construct the concept map test or the multiple choice test?

APPENDIX 5

Procedure for training students

Concept Mapping Training

A. Describe the scope of the study.

Say, "The purpose of this study is to examine the reliability and validity of a testing procedure as a way to assess what students know prior to instruction. An instructor would use this information NOT for grading students, but for planning instruction. If instructors know what their students know before a course begins, instructors may develop more meaningful instruction for their students. There are many ways to assess what students know prior to a course. For example, many instructors use a multiple-choice test. This study is looking at an alternative way to assess knowledge. It is called concept mapping. When producing a concept map, you describe relationships you understand to exist between concepts. This tells the instructor what you know about the key concept being mapped. The instructor then can use that information to plan appropriate instruction. For this study, you will be asked to produce a concept map on a topic in industrial engineering ("regression" was used for Study I) and then you will be asked to take a multiple-choice test on the same topic. The data will not be used as part of your course grade."

B. Consent Form

Review the "how involved" section with the class. Ask them to read the rest of the form and sign them if they are willing to participate. Collect consent forms.

C. Introduce the term "concept."

1. Make a list of words on the overhead using a list of familiar words for objects and events. For example, car, dog, chair, tree, cloud, book, raining, playing, washing, thinking, thunder, and birthday party.

2. Ask the students to describe what they think of when they hear the word car, dog, etc. Ask for volunteers to describe what they think of when they see or hear the word "car". Say, "Even though we use the same words, each of us may think of something a little different. These mental images we have for words are our concepts." Introduce the word "concept" and define as "a regularity in events or objects designated by some label" (Novak & Gowin, 1993, p. 4-5). The words listed here are our labels representing a regularity in events or objects. These words are called "concept words." . Say, "The differences in our mental images partially explain that our concepts are never quite identical even though we know the same words. Words are labels for concepts, but each of us must acquire our own meanings for words."

3. Say, "There are many related concepts within these concept words. For example, within the concept car may include engine, tires, and seat belt. There are related concepts between these concepts as well. For example, within engine may be fan belt, radiator, carburetor, and spark plugs. There is a hierarchical order to these concepts - from most general (e.g., car) to most specific (e.g., spark plugs)."

4. Show the concept word "engineer" ("graduate school" was used for Study I) on the overhead. Ask students to think back to when they were in elementary school and think about what this concept meant to them at that

time; then to think back to when they were freshmen in high school and what this concept meant to them at that time; then to think back to when they were seniors in high school and what the concept meant to them at that time; then to think about what the concept means to them today. Help students see that meanings of concepts are not rigid and fixed, but can grow and change as we learn and gain more experience.

5. List words such as are, where, is, the, with. Say, "What kinds of images come to mind when you hear these kinds of words? Probably not the same kinds of images as the concept words. These are not concept words, but linking words, and we use them in speaking and writing. Linking words are used together with concept words to construct sentences that have meaning."

6. Using two concept words and linking word(s), construct a few short sentences on the overhead to illustrate how concept words plus linking words are used by people to convey meanings. (engineers solve problems, learning is a part of growing up). Say, "This is where concept mapping comes in. If students can use linking words to demonstrate accurate relationships between concept words, then the instructor has some indication of what the students know."

7. Say, "We've said our images we have for words are our concepts, concepts change over time due to learning and experience, key concepts are made up of related concepts, and we can use linking words to demonstrate meaningful relationships between concepts. Therefore, through concept mapping we demonstrate our *current* understanding of a key concept by first arranging related concepts in a hierarchical order under the key

concept and then describing the relationships among the concepts using linking words." Ask if the students have any questions..

D. Demonstrate concept mapping.

1. Show examples of concept maps to the students. Point out the concepts and linking words in the maps. Demonstrate the hierarchical order from most general concept to most specific. Use the following examples (from Novak and Gowin, 1993):

   Plants

   Drinking water

   Concept maps (pass out the hard copy for them to keep as a reference)

2. Put up the "Example concept map" overhead. Tell them you will demonstrate how to build a concept map of the concept "golf" to demonstrate the process they will go through to build their own maps. Reinforce the following points:

   * If they don't understand how a concept word fits, then don't use it.
   * Describe the relationship between two concepts. One way the map helps the instructor to know what the student knows is by showing what relationship the student understands to exist between two concepts.
   * It is O.K. not to understand a concept. You are not expected to have a complete understanding of the topic you will map. But if you are somewhat familiar with these concepts it is important for the

professor to find out so that either misconceptions can be addressed in the course or emphasis can be added or taken away for specific topics.

Finish the example map by saying, "Most first effort maps are sometimes messy and you may not be completely happy with it. Consider what you have done and reconstruct the map if this is helpful. One and sometimes two reconstructions of a map may be needed to show a good representation of propositional meanings as they understand them."

E. Practice concept mapping

Pass out practice materials. Ask students to construct one more concept map using "VPI library" as the topic. Give feedback as they work on the maps.

APPENDIX 6

Questionnaire for Content Validity Data

Estimation of Content Validity

The purpose of content validity is to assess whether the items on the multiple choice and concept mapping tests adequately represent the domain of measurement from an ISE perspective, as presented in the Introduction to ISE course.

The domain of measurement from an ISE perspective is defined by the student performance objectives:
The student will be able to

1. Define work measurement, organizational performance measurement, and standard time.
2. Define effectiveness, efficiency, and productivity.
3. Explain the relationship between work measurement and organizational performance measurement.
4. Explain the relationship between effectiveness, efficiency, and productivity.

(Assume all objectives are weighted equally.)

A. Consider the following questions regarding the student performance objectives:

1. Please identify any objectives that go beyond the domain of measurement (from an ISE perspective, as presented in the Introduction to ISE course).

Expert A - None

Expert B - None

2. Please add any objectives that you think should be added to fully represent the domain of measurement (from an ISE perspective, as presented in the Introduction to ISE course). Use the back of this page if necessary.

Expert A - None

Expert B - Other objectives have already been identified by instructor (#5-12 on his lecture notes; however, 1-4 (above) are sufficient.

B. In the matrix, indicate how well each test item matches the objectives in terms of characteristics shown in the top row. These characteristics have been defined below. Use the following scale when making your judgment:

1 = poor match 2 = fair match 3 = good match 4 = excellent match

subject matter - Does the content of the item match the content of the objective?
level of complexity of the performance required - Is the required complexity of performance as indicated by the objective similar to that required by the test item?
format of presentation - Does the question type require the respondent to display a behavior beyond that required by the objective (e.g., does the question use vocabulary at a level beyond what the respondent is expected to have)?
mode of response - Does the nature of the response require skills beyond that required by the objective (e.g., does an answering the question require a mathematical skill beyond those specified or implied in the objectives)?

multiple choice test item number:	subject matter	level of complexity of the performance required	format of presentation	mode of responses
1	3.5	4.0	3.5	4.0
2	3.5	3.5	3.5	4.0
3	4.0	4.0	4.0	4.0
4	4.0	4.0	4.0	4.0
5	2.5	3.5	3.0	4.0
6	3.5	3.5	4.0	4.0
7	2.5	3.0	3.5	4.0
8	4.0	4.0	4.0	4.0
Average	3.5	3.7	3.7	4.0

concept mapping item	subject matter	level of complexity of the performance required	format of presentation	mode of responses
work measurement	4.0	4.0	4.0	4.0
organizational performance measurement	4.0	4.0	4.0	4.0
organizational system	4.0	3.5	4.0	3.5
efficiency	4.0	4.0	4.0	4.0
effectiveness	4.0	4.0	4.0	4.0
productivity	4.0	4.0	4.0	4.0
standard time	4.0	4.0	4.0	4.0
inputs	4.0	3.5	3.5	4.0
outputs	4.0	3.5	3.5	4.0
processes	3.5	3.5	3.5	4.0
tasks	3.5	4.0	4.0	4.0
Average	3.9	3.8	3.8	3.9

C. Overall, is there a balance between the emphasis on subcatagories (i.e., those concepts related to measurement from an ISE perspective) within the objectives and the emphasis on subcatagories as reflected by test items?

Expert A - Seems to be balanced

Expert B - Yes

APPENDIX 7

Expert Interview Data

Expert Interview Data

1. **Did the concept mapping procedure measure the concepts it was supposed to measure? Why or why not? How or how not?**

   Expert B - "Yes, concept mapping showed things the multiple-choice test can't show and the multiple-choice test showed things the cm test can't show."

   Expert A - usually "more procedure problem when measure didn't work." "Concept mapping skills were poor and may have made a difference. If they had had better concept mapping skills it may have been a more effective exercise."

   Expert B - mapping ability affected the results. Also, "some students did not work hard enough on the test - just did it to fill time - knew they would not be graded so didn't try very hard to do it right."

   Expert B - Concept mapping procedure is active, not inactive - it engages students. Helps students identify what they do and don't know. Kind of like a short answer test where you are asked to compare and contrast something - students are more active participants.

2. **Which do you prefer as a measure of what your students know (for the purpose of instructional planning) - concept maps or multiple-choice tests? Why?**

   Expert B - Concept mapping - not for student evaluation (grading) but for planning, yes. Can see knowledge structures of students and how things fit together. "(Students) can't mimic with concept mapping - they must understand the material you can memorize and do well on a multiple-choice test, but not with a concept mapping test."

   Expert A- Concept mapping "easier to construct a concept mapping exercise" writing multiple-choice questions is time consuming and difficult. "The multiple-choice test question is either right or wrong, can't get much more than that like why student answered the question wrong." The concept mapping provides richer data. One negative for the concept mapping - students have to be taught how to concept map and have to have some proficiency in mapping.

3. **In what ways was the information your students provided on the concept map different from the information they provided on the multiple-choice test?**

   Expert B - "Concept mapping showed knowledge structure, how do the pieces fit together - shows the "big picture". Concept mapping assess student knowledge structures. Multiple-choice can't - can get at the relationship between two terms, but not the big picture - small pieces with the multiple-choice, versus the big picture with the concept map.

   Expert A - with multiple-choice each question is independent, can't see how they (students) fill in the gaps.

4. **Do you feel confident that your students' knowledge of measurement (from an I.E. perspective) was fairly assessed using the concept mapping procedure? Why or why not?**

   Expert B - parts of what they know about measurement not shown on the map - maybe a problem with the concept mapping skill.

   Expert A - depends on effort they want to put into it. "Multiple-choice is not as dependent on student effort."

5. **Do you feel confident that your students' knowledge of measurement (from an I.E. perspective) was fairly assessed using the multiple choice test? Why or why not?**

   Expert A - "for limited objectives - didn't go a step beyond." We did eliminate one question.

   Expert B - yes, (agreed with Expert A)

6. **Would you use the concept mapping procedure as tool for instructional planning?**

   Expert B - sees potential as an evaluation tool, but not as part of students' grades

   Expert A - "good in-class exercise - good teaching tool or use as a discussion tool."

Expert B - "yes, would use before a lecture, especially for classes where they know something, but coming from different perspective (many different majors in a class). If they (students) know nothing about the topic - probably wouldn't work."

Expert A - "yes, under the right conditions (as Expert B described) particularly with grad students and undergrad students in the class" - different levels of knowledge and experience. "Could see application in training - assess learners prior to developing training."

7. **Were you able to identify anything from the concept mapping pretests that would suggest areas to target during instruction?**

Expert A - "Sure - there was confusion between definitions and relationships between efficiency, effectiveness and productivity Also connection between organizational performance measure and work measurement was not apparent."

Expert B - "efficiency, effectiveness, and productivity showed confusion here. Some had good hierarchies, some had wrong things on top. For example, work measurement was towards the top, over organizational system when this should have been reversed. Students with more experience probably did better."

8. **Do you think a scoring procedure is necessary in order for you to make decisions for instructional planning?**

Expert B - "useful, but not necessary - still got plenty of information just by ranking."

Expert A - agree - "There is value from just looking. A scoring procedure would help to aggregate the data more objectively." Could look at the best and worst maps and see why they are best and worse. From scores, ordinarily would rank, then look at how tests clustered and why. Have to look at sets of maps to make instructional decisions.

APPENDIX 8

Student Questionnaire Data

Student Questionnaire Data

1. Which do you prefer for measuring what you know - concept mapping or multiple-choice tests? Why?

#1
Multiple choice tests. Because there are choices. If you are completely lost, the concept mapping is going to be extremely difficult.

#2
Concept mapping. It makes you think more because you have to show items in relation to each other in the hierarchy.

#3
Concept mapping, because it really makes you think about the relationships between concepts. With multiple choice, if you don't know much about the concepts, I might guess rather than really thinking about it.

#4
Multiple choice - because it is a more definite way of getting to an answer. Concept mapping cannot be applied to everything.

#5
Multiple choice, it is less subjective.

#6
Multiple choice questions, because there is a limit to the number of answers. You simply pick the best answer. Concept maps can be drawn in many different ways and even though there is only one best way, the different possibilities can confuse me. Sometimes multiple choice questions are unfair as in they are purposely written in a confusing manner. This does not test a person's knowledge of the subject matter.

#7
The concept map is more logical and more challenging - I like it because it shows more what I am thinking not what I have memorized. It's a little more risky to do a concept map - but I like it.

#8
I preferred the multiple-choice question simply because the questions provided clues to help sort my thought [sic] and come up with a correct answer. However, the concept mapping required deeper thought and only could tap deep seated understanding of the topic, whereas the multiple choice could stir vague recollections which made the test easier.

#9
I prefer multiple-choice tests. Because, generally, there is only one "correct" answer in a multiple choice question. If I know the answer, it's quite easy to select the correct response One the concept mapping tests, there are too many ways to present the answer and it's hard to tell what would be considered the correct response.

#10
Concept mapping can demonstrate the knowledge of the student much better than any guessing (multiple choice) test can measure. Although it is much more subjective to the instructor opinions.

#11
Multiple-choice - the concept map is a little more complicated and there is more than one answer there are not exact answers.

#12
A 70-30 split. 1. concept mapping allows freedom of expression. Allows logical statements to be presented to back up choices. If a multiple choice question is ambiguous and misinterpreted a different and therefore wrong answer will be selected, even if you know the right answer. 2. multiple-choice provide memory aids, in that you might recognize the right answer if you see it.

#13
Concept mapping - it shows the relationships of elements, it is similar to [sic] cause-effect diagram, it is a way of system thinking, it help [sic] us to memorize the information.

#14
I prefer multiple-choice, though I think that may only be because I am more comfortable with that method. Perhaps with more exposure to mapping I would prefer that method because it doesn't so much imply an all right or all wrong answer it allows for more flexibility in response.

#15
Concept Mapping. Sometimes you may know what you want to say, but with a multiple choice test you may not find the answer that you are looking for. Concept mapping allows you to state exactly what you think something means.

#16
Concept mapping gives a better feel for how well a student knows the topic. Multiple choice only gives information if the student knows a specific question.

#17
Well, the multiple-choice were [sic] than your test's multiple-choice questions. So, I'm not sure they measure may knowledge on the topic. The concept map affords me the ability to express relationship between concepts - the concept map I would say I prefer.

2. How was the information you provided on the concept map different from the information you provided on the multiple-choice test?

#1
The info on the mapping was much more vague. Not a lot of detail.

#2
The concept map showed more of what I knew about the subject because of the relations.

#3
Much more thought went into the concept map.

#4
Concept map was more vague and I'm sure someone else set it up in a different way, but not in the multiple choice.

#5
Different tasks, different instructions.

#6
blank

#7
Concept map - understanding in terms w/out guide for meaning
multiple choice - choices may give a direction because they may contain key words to trigger memory of meaning.

#8
Again, the questions in the multiple choice section provided memory stirring [sic] clues which helped me in remembering what I had learned.

#9
More thinking and organization skills required. Sometimes it also depend on "creative" thinking of how to link all concepts on the problems in a cohesive fashion.

#10
Multiple choice tests don't exercise the student critical thinking ability. The concept maps show interrelationships between items that could not be covered by a multiple-choice test.

#11
I used the answers in the multiple choice part to help with the concept map.

#12
Allows me to provide reasoning, allow the exact titles, names, etc. may not be correct but the understanding of the "concept" presented might be right.

#13
In multiple-choice: the information is random and mixed.
In concept map: the information is much complete, we just organized them into the system picture.

#14
It forced me define relationships instead of just deciding which of the given choices sounded right.

#15
The difference is that I am not sure If I selected the "best" answer on the m.c. test. but with the concept mapping technique I know that I am at least on the right track.

#16
The concept map gives a relationship between all the terms instead of separating them all into separate questions.

#17

Concept map allowed me to express relationships.

3. Do you feel confident that your current knowledge of measurement (from an industrial engineering perspective was fairly assessed using the concept mapping procedure? Why or why not?

#1
Yes, but if you wanted specifics (and I knew them) then it wouldn't be a good test.

#2
Yes, if I was not knowledgeable it would probably show pretty easily.

#3
Yes, between the lectures and the text, the relationships between these concept was spelled out.

#4
No, some of the concepts fit in more than one place but was difficult to present.

#5
I don't really know because I don't know if I was right or wrong on the concept mapping procedure. Give feedback and then I can answer.

#6
No, because I am still not sure about mapping. I don't think I could map out something I know very well right know [sic] and give as much info as it was to write an essay.

#7
There is alot [sic] of room to miss information that the person grading the test is looking for. I think there needs to be more guidelines about how specific you need to be in this case.

#8
No, I couldn't demonstrate some of the relationships I understand.

#9
No! I feel that I could provide a dozen more maps of my knowledge of measurement, depend [sic] on which perspective I want to present it.

#10
No. The concept maps should be complemented with an essay based exam, explaining the process or details for the concept map that can't be shown.

#11
No, I don't think that the concept map gives a good assessment of one's knowledge in any subject. No 1 test can give an [sic] complete assessment of one's knowledge in a subject. The more types of tests the better.

#12
yes

#13
No. Since the measurement is only part of (or subsystem) of an industrial engineering, we need more knowledge to link the measurement field with the other fields of IE.

#14
Yes, for the most part it did, keeping in mind the fact that different people see things in different ways so a different answer isn't necessarily wrong.

#15
No. I could have probably went into more detail had I taken a different kind of test, e.g. essay test.

#16
Yes, the concept map gave a broader perspective on how well I know the topic.

#17
Maybe! I feel it could be a good measure; however, I'm not sure I had the motivation to do well.

4. Do you feel confident that your knowledge of measurement (from an industrial engineering perspective) was fairly assessed using the multiple-choice test? Why or why not?

#1
Yes but if you wanted just general ideas the multiple choice test was too specific but it could have been made as general as concept mapping with different questions.

#2
I am unsure because some answers were confusing.

#3
No. Some of the differences between answers seemed small, a level of detail we didn't have time to go over in class. Seemed more memorization than knowledge based.

#4
Yes, but maybe more questions could've been asked to cover all areas.

#5
see answer to question 3

#6
Yes, but I do believe [sic]

#7
More so because you know exactly what a grader wants - it is more cut and dry.

#8
No, due to the clues I received it was able to help me.

#9
More so than the concept mapping. For reasons [sic] that I would need to have sufficient knowledge about measurement in order to answer any kind or type of multiple-choice questions correctly.

#10
I've never thought that a multiple choice test fairly assessed anybody's knowledge.

#11
No. see quest 3.

#12
No - no problem solving, no creative processes

#13
No. Using the multiple-choice test, only pieces of information are shown, through multiple-choice test, I still have no confidence that I know the whole concepts of measurement. Besides, in the multiple-choice test, I can guess the "right" answer, but it doesn't mean that I understand the measurement.

#14
Yes, but it allows a lucky guess to be mistaken for concrete knowledge.

#15
No. Because some of my answers may be incorrect.

#16
No, the multiple choice is very specific and does not give an overall perspective of how well I know the topic.

#17
Not at all. But I feel that it is a combination of multiple-choice test in general, and that your test was to brief and leading from the material in class.

5. About how long did you work on your concept map of measurement?

#1
5 minutes

#2
7-8 minutes

#3
10 minutes - Question, do most people tend to round off to the nearest 5 min like me?

#4
about 10-12 minutes

#5
5 min

#6
15 minutes, more time couldn't have helped anyway

#7
3/4 of time given

#8
15-20 min.

#9
5-10 min.

#10
7-10 minutes

#11
10 minutes

#12
1st time - it seemed like 30 minutes because it was grappling with the ideas and concepts. I did not have a full understanding of.
2nd time - about 10 minutes. I understood measurement better.

#13
15 min.

#14
10 minutes

#15
about 10 minutes

#16
10 min

#17
12 min

6. About how long did you wok on your multiple-choice test of measurement?

#1
3 minutes

#2
10-1 minutes

#3
5 minutes

#4
about 5 minutes

#5
3 min

#6
5 minutes

#7
1/4 time given

#8
10 min.

#9
1 min to 2.

#10
2-3 minutes

#11
10 minutes

#12
1st time - 5 minutes max
2nd time - 3 minutes max

#13
5 min.

#14
5 minutes

#15
about 6 minutes

#16
5 min

#17
6 min

7. In general, what do you think about concept mapping as a way for an instructor to assess what you know prior to instruction with the intention of using this information to plan instruction (NOT to assign grades to students)?

#1
The concept mapping is helpful in finding out generally what a student knows but not specifics.

#2
I think it is a good way to evaluate the knowledge of the student.

#3
Good idea - plus it gives students a forewarning of what the instructor thinks is important when that topic is covered.

#4
No, because the student may just try to fit everything in without a real idea of the concept.

#5
It would probably be a better gage than multiple choice because guessing is not a factor.

#6
It would be good if I was more practiced at mapping, however, I do not think it could give any more awareness than a well thought out (as in well written) multiple choice test, when trying to discover a students knowledge of new material, strike that. I think it would be a much better way to gain insight on a students knowledge.

#7
It's okay because it lets the instructor know how a student views that subject - maybe s/he needs to give the student a direction that is more appropriate. It could help a student if used effectively.

#8
An excellent way. It would provide the instructor with a general level of understanding that the class has mastered to that point.

#9
blank

#10
Its a good idea, but like I mentioned before it should be complemented with an essay type test, this would help in explaining the concept map.

#11
I'm sot sure it is a good assessment.

#12
It points out glaring deficiencies, erroneous concepts as well as strengths and weaknesses. It would work better for a straight freshman class. The background of this particular class is too varied.

#13
It is good for both instructor and student.

#14
I think it could be constructive.

#15
I think it is an okay idea.

#16
I think it is a fairly good way to assess knowledge.

#17

I seems to me that it's more important to plan instruction based on the whole curriculum of ISE and on entry-level needs of students. Provide the basics regardless of student knowledge as long as you don't instruct at a level beyond entry-level capabilities. Listing the concepts (linear) for a hierarchical (hyper) map does not make the cognitive mappings between the two steps compatible. List them in groups. However, concept mapping I believe is useful for assessing knowledge, not planning instruction.

SUSAN L. COLEMAN

EDUCATION

Ph.D. Instructional Technology and Design. Virginia Polytechnic Institute and State University, 1994. Dissertation title: Estimating the reliability and validity of concept mapping as a tool to assess prior knowledge. Chair: Thomas Sherman.

M.A. Curriculum and Instruction with emphasis on instructional design. Virginia Polytechnic Institute and State University, 1990.

B.A. Education (concentration in science). Virginia Polytechnic Institute and State University, 1983.

PROFESSIONAL EXPERIENCE

1993-1994 Senior Research Associate, Instructional Technologist - Management Systems Laboratories, Total Quality Management Training Group. Execute all phases of instructional systems design on multiple, simultaneous projects within the technical and non-technical training arenas, including job task analysis, training needs analysis, and program evaluation. Primary projects include emergency management training, a technical intern program, software training, behavior-based safety training, mentor training, and train the trainer sessions. Pursue research related to instructional technology, design, and training. Author and co-author papers in the area of instructional technology and design. Mentor co-workers who have interests in the field of instructional design. Serve on continuous process improvement team.

1991-1993 Research Associate, Instructional Technologist - Management Systems Laboratories, Training Systems Laboratory. Project manager for FY93 Department of Energy Applied Engineering Fundamentals course. Collaborated on a daily basis with the lead professor and Department of Energy (DOE) sponsor regarding course modifications and development of new course material. Provided instructional consulting services to VPI engineering faculty and Duke Engineering & Services practicing engineers during development and implementation of the course. Coordinated logistics, purchasing, and faculty payroll with the Director of Program Development at the VPI Continuing Education Center. Created guidelines for the development of instructionally sound case studies. Reviewed all course materials to ensure adherence to quality assurance standards adopted by DOE. Delegated assignments to two Information Officers and two Office Services Specialists supporting the course. Updated Program Director of course progress. Worked with an evaluation team to design, develop, and implement data collection procedures for the course evaluation. Interpreted and summarized selected portions of the evaluation data. Authored and co-authored papers in instructional technology, training, problem solving , and instructional design.

1991	Summer Intern - Management Systems Laboratories, Technology Transfer Group. Established a process for VPI faculty to create lesson plans for modules comprising the FY92 Applied Engineering Fundamentals Course. Edited lesson plans being created for each module. Provided guidance to subject matter experts developing performance-based objectives. Researched the areas of problem solving, effective application of case study materials, and engineering education for a grant with DOE.
1990-1991	Graduate Research Assistant, Residential Success Program, Virginia Polytechnic Institute and State University. The Residential Success Program was a joint research project between the College of Education and the Office of Housing and Resident Life. The purpose of the program was to provide effective and comprehensive retention services to high-risk students. Designed and delivered an introduction to computers seminar. Worked individually with freshman students to identify skill deficiencies and provided or recommended corrective actions. Supported the principal investigator in conducting weekly study skills seminars. Gathered and analyzed data concerning study behaviors of first-year college students. Administered questionnaires to survey study skill behaviors and assess students' adaptation to college life.
1986-1990	Administrative Support, The Center for Advanced Ceramic Materials, Biology Department, and the College of Architecture, Virginia Polytechnic Institute and State University. Within the Ceramics Center, functioned as liaison between the Community College Economic and Development Directors and our Center, working to transfer technology from the Center to Virginia businesses involved in ceramics. Wrote articles for the Center's quarterly newsletter and assisted with layout design. In the academic offices, registered students for classes and helped them change majors, interpret graduation analyses, and select an advisor. Frequently interacted with other offices throughout the University, including Dean's offices, Department offices, the Financial Aid Office, Student Accounts, and the Registrar. Developed a PC users manual, then designed and delivered a training program addressing the operation of the IBM PS-2 for the Biology office staff.
1985-1986	Sixth-Grade Teacher, Boone County Public Schools, West Virginia. Taught all subjects to a sixth grade class in a self-contained classroom. Participated in selection of new math curriculum.

1984-1985 Substitute Teacher, Kanawha County Public Schools, West Virginia. Taught from Kindergarten through the sixth grade, including special education, physical education and library classes. Substituted more than 130 days through the school year.

MAJOR TRAINING AND DEVELOPMENT PROJECTS

Emergency Management Training - a job task analysis for Emergency Operations Center Staff positions in Fermco's Emergency Preparedness Department at the Fernald environmental restoration site. The job task analysis focused on the exceptional tasks and responsibilities required for responding to site emergencies. Results will be used to update and enhance the emergency preparedness training program and make recommendations for re-alignment of emergency duties among positions.

Applied Engineering Fundamentals Course - a 14-week multidisciplinary course developed for the U.S. Department Of Energy's Office of Defense Programs and Office of New Production Reactors. Participants were newly hired engineering interns from a variety of engineering backgrounds and with little or no prior relevant work experience . The purpose of the course was to educate the interns in the area of nuclear engineering so they can eventually assume oversight responsibilities at a DOE nuclear facility. A variety of instructional methods were incorporated into the instructional strategy including case studies, site visits, and peer tutoring. An extensive evaluation was conducted of the course, as well as a follow-up evaluation one year later.

Technical Intern Program - a three year program developed for the U.S. Department of Energy's Office of Defense Programs. Participants were the engineering interns referred to above. Interns began the program with an Orientation to DOE and Defense Programs. Then, interns attended the Applied Engineering Fundamentals Course in Blacksburg, Virginia to build a solid technical knowledge base. After the course, interns rotated to different DOE facilities. During each 6-month rotation, objectives and a plan for meeting those objectives were developed for each intern. Evaluation of the Technical Intern Program was conducted each year.

Alpha Four Software Training - a training program developed for the U.S. Department of Energy's Office of Environmental Restoration and Waste Management (EM). Participants were employees of EM, some having a great deal of experience with computers and some having little or no experience. The purpose of the course was to introduce participants to the software package and show them how to set up a simple database, enter data, and create summaries and reports. The Alpha Four training program consisted of two components - a classroom training component, and an independent learning component consisting of video-taped instruction and accompanying workbook.

Behavior-Based Safety Training - training developed for organizations improving the effectiveness of their existing industrial safety programs. The purpose of this program is to train employees to properly implement tools that increase safe behaviors and decrease unsafe behaviors in their workplace. In addition to providing direct training, train-the-trainer courses were developed. Modules for the train-the-trainer course were constructed so trainers could easily customize the delivery to site-specific needs. Customers include the Westinghouse Hanford Company (WHC) and the Idaho National Engineering Lab (INEL). WHC participants included employees from operations and maintenance to high-level management positions. In addition to training over 3200 WHC employees, selected WHC employees were trained to deliver behavior-based safety training for the remaining 10,000 plus employees. At INEL, trainers were trained to deliver the behavior-based safety training.

Empowering Employees Through Successful Coaching - the training component of a larger program to introduce mentors into a work environment with a poor safety record. The mentors were tasked with upgrading the conduct of operations, focusing primarily on safety concerns. Participants were mentors, supervisors, and protégés (operators) in the Steam and Water Utilities Division of WHC. The purpose of the training program was to prepare the mentors for their role as mentors, provide background information to mentors and supervisors regarding reasons why people perform unsafe behaviors, and guide mentors and supervisors in determining solutions to implement in the workplace. Interpersonal communication skills, role clarification, and the qualities and behaviors of a good mentor or coach were addressed in the mentor orientation workshop. The remaining workshops addressed six factors influencing the performance of unsafe acts, learned helplessness, perceived reward, perceived risk, training, task demands, and equipment and environment design.

PROFESSIONAL DEVELOPMENT

Attended "Continuous Improvement Training Workshop", taught by Dr. Brian Kleiner of Management Systems Laboratories, at Virginia Polytechnic Institute and State University, September, 1993.

Attended "Preparing The Future Professoriate", a graduate student/new faculty seminar conducted by the Office of Sponsored Programs, at Virginia Polytechnic Institute and State University, January 7-8, 1993.

Attended "Self-Perception and Management Values" course (one day) taught by Leo McManus of McManus Associates, in Blacksburg, Virginia, June 5, 1992.

Attended "The Contracts and Grants Process", a graduate student topical seminar taught by the College of Education at Virginia Polytechnic Institute and State University, April 20, 1991.

Participated in the logistical planning and implementation of the Virginia Polytechnic Institute and State University teleconference "Making Sense of TQM", presented by Dr. D. Scott Sink and Dr. Ralph Badinelli, March 7, 1991.

Attended an introduction to the Myers-Briggs Type Indicator taught by Delta Smith of the Counseling Center, Virginia Polytechnic Institute and State University, November 1989.

AFFILIATIONS

American Society for Engineering Education
National Society for Performance and Instruction
American Educational Research Association
Eastern Educational Research Association

AWARDS AND HONORS

Dean's List
Pi Society of Pi Beta Phi
Instructional Fee Scholarship from the College of Education, Division of Curriculum and Instruction:
- Fall Semester, 1990
- Fall Semester, 1991
- Spring Semester, 1992
- Fall Semester, 1992
- Spring Semester, 1993

PROFESSIONAL CONTRIBUTIONS

Publications

Coleman, S. L., & Tomchin, E. M. (1993). An instructional strategy to integrate topics within a multidisciplinary course. In J.P. Moshen (Ed.) *Proceedings of the American Society of Engineering Education Southeast Section* (pp. 358-365). Washington, DC: American Society for Engineering Education.

Coleman, S. L., Coleman, G. D., & Johnston, C. S. (1993). Managing and measuring a training function: Applying TQM principles. In D. J. Sumanth, J. A. Edosomwan, R. Poupart, & D. S. Sink (Eds.), *Productivity & Quality Management Frontiers - IV*, (Vol. 2, pp 942-953). Norcross, Georgia: Industrial Engineering and Management Press.

Technical Reports

Coleman, S. L. & Hughes, J. (1994). *Final report: Job task analysis for Fermco emergency operations staff.* Virginia Tech. Final report for a contract sponsored by the Fernald Environmental Restoration Management Corporation.

Casali, S., Coleman, S., Welch, K., Carolan, T. & Kotnour, T. (1993). *Why do people perform unsafe behaviors? A review of the scientific literature concerning underlying factors affecting the performance of unsafe acts in the workplace* (Report No. MSJ-SRV-287290). Richland, Washington: Westinghouse Hanford Company.

Koball, E., Coleman, S. L., & McClintock, M. (1993). *DP technical intern program end of year one evaluation.* Virginia Tech. Program evaluation for a grant sponsored by the U.S. Department of Energy.

Tomchin, E. M. & Coleman, S. L. (1992). *93-01 Applied engineering fundamentals course evaluation.* Virginia Tech. Course evaluation for a grant sponsored by the U.S. Department of Energy.

Tomchin, E. M., Murawski, M. N., & Coleman, S. L. (1991). *92-01 Applied engineering fundamentals course evaluation.* Virginia Tech. Course evaluation for a grant sponsored by the U.S. Department of Energy.

Presentations

Bilger, W. T., Shropshire, J. C., Coleman, S. L., and Collins, R. (1993, November). *Technical leadership development: Speeding up the process.* Paper presented at the annual meeting of the American Nuclear Society, San Francisco, California.

Coleman, S. L., & Tomchin, E. M. (1993, April). *An instructional strategy to integrate topics within a multidisciplinary course.* Paper presented at the annual meeting of the American Society for Engineering Education Southeastern Section, Nashville, Tennessee.

Murawski, M. N., & Coleman, S. L. (1993, April). *Integrating content and context: An instructional strategy to simulate the work environment.* Presented at the annual meeting of the National Society for Performance and Instruction, Chicago, Illinois.

Coleman, S. L., Coleman, G. D., & Johnston, C. S. (1993, February). *Managing and Measuring a Training Function: Applying TQM Principles.* Paper presented at the Fourth International Conference on Productivity and Quality Research, Miami, Florida.

Coleman, S. L., & Sherman, T. M. (1992, March). *Concept mapping: A tool for knowledge assessment.* Presented at the annual meeting of the Eastern Educational Research Association, Hilton Head, South Carolina.

Coleman, S. L., & Sherman, T. M. (1992, April). *Concept mapping: A tool for knowledge assessment.* Poster presentation at the Virginia Tech Graduate Student Assembly Graduate Research Symposium, Blacksburg, Virginia.

Susan L. Coleman