From **Gal, I. & Garfield, J. B. (editors).** **The Assessment Challenge in Statistics Education**. IOS Press, 1997 (on behalf of the ISI). Pages 91-104. ISBN 90 5199 333 1. The References section is a separate chapter and is available on the Internet at http://www.stat.auckland.ac.nz/~iase/publications/assessbkref.

Chapter 8

Assessing Students' Connected Understanding of Statistical Relationships

Candace Schau and Nancy Mattern

Purpose

We believe that connected understanding among concepts is necessary for successful statistical reasoning and problem solving. Two of our major instructional goals in teaching statistics at any level are to assist students in gaining connected understanding and to assess their understanding. In this chapter, we will explore the following questions:

- Why is connected understanding important in statistics education?
- What models of connected understanding are useful in thinking about statistics learning?
- How can connected understanding be represented visually?
- What approaches exist for assessing connected understanding?

INTRODUCTION

We, and many other statistics instructors, routinely observe a critical weakness in post-secondary students who have taken applied statistics courses: they lack understanding of the connections among the important, functional concepts in the discipline. Without understanding these connections, students cannot effectively and efficiently engage in statistical reasoning and problem-solving. They remain novices. They have "isolated" knowledge about various concepts; for example, they may be able to calculate a standard deviation and a standard error. However, they do not understand how these concepts are related (and distinguished) and so make application mistakes such as using one concept when they should have used the other. Some students recognize their lack of connected understanding and will say things like "I can solve a problem using the t-test when I know I'm supposed to. But otherwise I don't have a clue."

There is a growing body of research with findings that attest to the statistical and probabilistic misconceptions held by students of all ages, as well as by adults. Garfield and Ahlgren (1988) cite study after study concluding that students cannot use statistical reasoning effectively in probabilistic situations nor can they solve statistical problems, even after exposure to instruction in statistics. We contend that these difficulties are due to students' lack of connected understanding of concepts.

The current work on national standards and goals in our country supports our belief in the importance of conceptual connections. The *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989) explicitly contains a standard for each set of grades called "Connections" (standard #4). This standard for Grades K-4 includes five desired outcomes such as "link ... knowledge" and "recognize relationships" (p. 32). For Grades 5-8, the five desired outcomes include, for example, "see mathematics as an integrated whole" (p. 84). For Grades 9-12, the four outcomes include, for example, "use and value the connections among mathematical topics" and "use and value the connections between mathematics and other disciplines" (p. 148).

To determine the extent to which both formative and summative instructional connection goals are met, instructors need assessment formats that measure connected understanding. This chapter describes our thinking about this kind of assessment. It is based on our work studying post-secondary students' connected understanding of statistics, as well as middle-school and post-secondary students' connected understanding of science.

MODELS OF CONNECTED UNDERSTANDING

There are many ways to think about learning. One fruitful way for us as teachers and researchers has been to think about the mental networks students form in the process of learning and use in applying their learning. The schema is an important concept in many theories of mental networks. Schemas and connected groups of schemas often are called cognitive structures, cognitive networks, or structural knowledge (Skemp, 1987).

A schema is a mental storage mechanism that is structured as a network of knowledge (Marshall, 1995). A schema results from repeated exposure to problem-solving situations that have features in common. The learner forms a schema by abstracting the most relevant of these features and either assimilating these features into existing schemas or creating new schemas. One critical defining feature of a schema is the presence of connections; in order to function, the components within a schema must be interconnected. As students (who are "novices") become more expert, their schemas gain components and these components and the schemas themselves become more interconnected; their cognitive structures begin to resemble those of their instructors (who are "experts"). Expertise is not merely knowing "a lot"; it is having a rich, accurate, and relevant set of interconnected schema and schema components (Marshall, 1995; Skemp, 1987). According to this model, then, students learn statistics through 1) assimilating (connecting) new information into their cognitive networks, 2) forming new connections among knowledge that already exists in their networks, 3) reorganizing (accommodating) their connected schema to match incoming information, and 4) eliminating incorrect concepts and connections.

Learners can (and often do) develop inaccurate schemas. Inaccurate schemas are most often formed when one or both of two conditions exist. The first condition reflects the state of the learner; the second concerns the problem-solving tasks learners experience (Skemp, 1987). First, all but the most basic schemas are built on existing schemas. Some students lack these prerequisite schemas. Others may have them, but they may be inaccurate; that is, some learners possess misconceptions when they enter statistics classes.

Second, problem-solving tasks may interfere with accurate schema formation in one of three common ways. Students may not experience enough tasks to be able to abstract the needed relevant common features. The tasks may lack these features. The tasks may have many features in common, and students may not be able to determine which are relevant and which are not.

Whether learners lack prerequisite schemas or problem-solving tasks are inadequate, resulting statistical schemas aren't developed at all, are incomplete, or are inaccurate. Both sets of situations require instructional planning and intervention. Statistics instructors need to identify and address important misconceptions before moving on to more complex knowledge. They also need to carefully develop (or select and evaluate) the problem-solving situations they use in instruction.

Schemas develop because they are useful in problem-solving. When faced with a task, problem-solvers first use their existing schemas to create a mental model of the problem. They then use schemas to plan an approach to solving the problem, to execute that approach, and to evaluate the outcome. The problem-solvers stop if the process was successful, try again until it is, or quit trying. To begin this process successfully, problem-solvers must recognize the critical features of the problem and match them to existing schemas; if students fail to form a mental model at this stage or if they form an incorrect mental model, the entire problem-solving process fails (Marshall, 1995; Skemp, 1987).

Statistics students can attempt to form mental models in at least two ways. First, they can try to abstract the important components of the problem and match these to their existing schemas. For the student to be successful, these schemas must be relevant and accurate. This approach is flexible and adaptable to many different problem-solving situations and tasks. Alternatively, if students do not possess relevant and accurate schemas or if they do not invoke them, they can attempt to form their mental models through trying to recall the appropriate formula. This approach becomes increasingly harder as more and more formulas are encountered in learning. Initially, it may be easier to memorize the formula needed to solve a particular problem than to form a schema, but eventually the sheer number of formulas and different problems becomes overwhelming for students' memories (Skemp, 1987). In addition, there are many problems (perhaps the most important ones) that cannot be directly solved by "plugging" into a single formula and "chugging" out the answer, either by hand or on a computer.

MAP REPRESENTATIONS OF CONNECTED UNDERSTANDING

Understanding in statistics, then, requires the existence and use of relevant and accurate schema that, by their nature, are organized in a mental structure. This organization can be represented as a visual-spatial network (or map) of connected concepts. Perhaps the most widely known, and inclusive, of map formats is the concept map (Novak & Gowin, 1984). A concept map includes the concepts (referred to as nodes and often represented visually by ovals or rectangles) and the connections (referred to as links and often represented with arrows) that relate them. In this representation, the basic unit is a pair of connected concepts called a proposition and consisting of two nodes connected by a link. Like schema networks, maps can, but do not need to, include hierarchical structures. See Holley and Dansereau (1984); Jonassen, Beissner, and Yacci (1993); and Novak and Gowin (1984).

Good concept maps include enough important concepts to represent the targeted subject area, and these concepts are clearly linked with important relationships. Figure 1, for example, presents our concept map representation of one possible hierarchical network that is relevant to us when we teach introductory statistics. In Figure 1, the proposition "statistics can be inferential" consists of the concepts of "statistics" and "inferential" connected by the link "can be." Propositions often are grouped into neighborhoods (clusters of propositions that are more densely related to each other than to other propositions). Figure 1 contains two neighborhoods, one containing propositions related to descriptive statistics and the other propositions related to inferential statistics. The inferential neighborhood contains two subneighborhoods, one related to "Variables that are Continuous" and the other to "Variables that are Discrete."

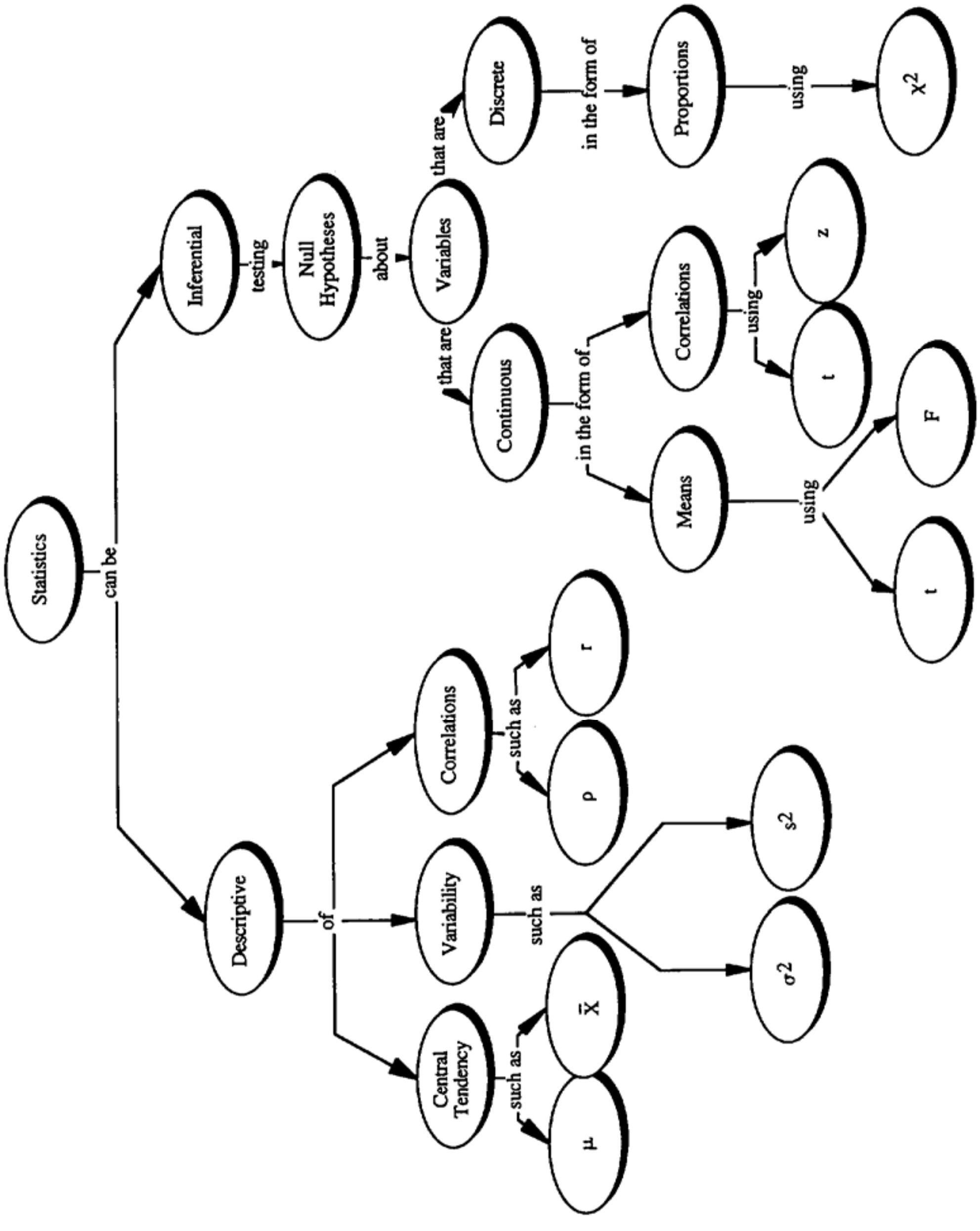

Figure 1. "Expert" (instructor-made) concept map representing aspects of one organization for an introductory statistics course

Figure 2 is a hierarchical map we created to represent the process of hypothesis testing of means. It contains two neighborhoods; the neighborhood under "Discrete independent variable(s)" contains two subneighborhoods. Neighborhoods may be cross-linked with each other; the "tested using" link connects neighborhoods 1 and 2 in this figure.

We construct "expert" maps like those in Figures 1 and 2 in several steps. We first identify the purpose that the map is to serve (an extended version of Figure 1, for example, was designed as a global map of course content). Second, we select the concept(s) that will serve as the foci for the map (in the case of Figure 1, the central concept is "Statistics"; second-level concepts are "Descriptive" and "Inferential"). Third, we identify the important concepts that are related to the central and second-level concepts and how these concepts are related. Fourth, we roughly sketch the map. Fifth, we draw the map using a software package called Inspiration (1994) or by hand. Sixth, we discuss our map and revise it, often by eliminating concepts, working on links, and rearranging propositions and neighborhoods.

Concept maps and other mapping techniques are used for at least three instructional purposes. First, they are used as an instructional planning tool. Teachers can use mapping in designing content coverage materials (such as handouts and notes), sequencing instructional delivery, and determining delivery strategies (such as activities and presentations). Creating these maps forces instructors to identify important connected concepts. It is impossible to draw maps without making explicit your implicit mental characterizations. In our experience, what we do not understand becomes painfully clear as we try to draw our maps. Second, maps are a learning tool (Harnisch, Sato, Zheng, Yamagi, & Connell, 1994). They may be prepared by a teacher and then shared with students with the intent of portraying a representation of the instructor's connected understanding of the material. Instructors also can assign map creation to students to encourage connected understanding and schema-building. There is a growing body of research indicating that both of these uses assist many students in learning. Third, these mapping formats can be and are used for both formative and summative assessment, either through map creation or map completion by students. Concept maps are the mapping format most widely used for assessment.

Our theory of how students learn should drive our development and use of appropriate assessments (Marshall, 1995). We believe that understanding requires connected concepts. If connected understanding is an important instructional goal, we must be able to assess it. In this chapter, we emphasize the use of concept maps for assessing connected understanding, contrasting the maps with assessments consisting of traditional item types.

USING TRADITIONAL ITEM TYPES TO ASSESS CONNECTED UNDERSTANDING

Two traditional item types—multiple choice items and word problems—have a long history of use for assessing achievement in statistics and mathematics. The common approach to the use of traditional assessment items includes the assumption (usually implicit) that understanding is an accumulation of bits of isolated information (Marshall, 1995). A test often consists of a collection of items measuring primarily recall.

Both of these item types can be used to measure aspects of connected understanding. Even when they are written to assess cognitive structures, they yield limited information: more information is obtained from analyzing errors than by analyzing correct answers.

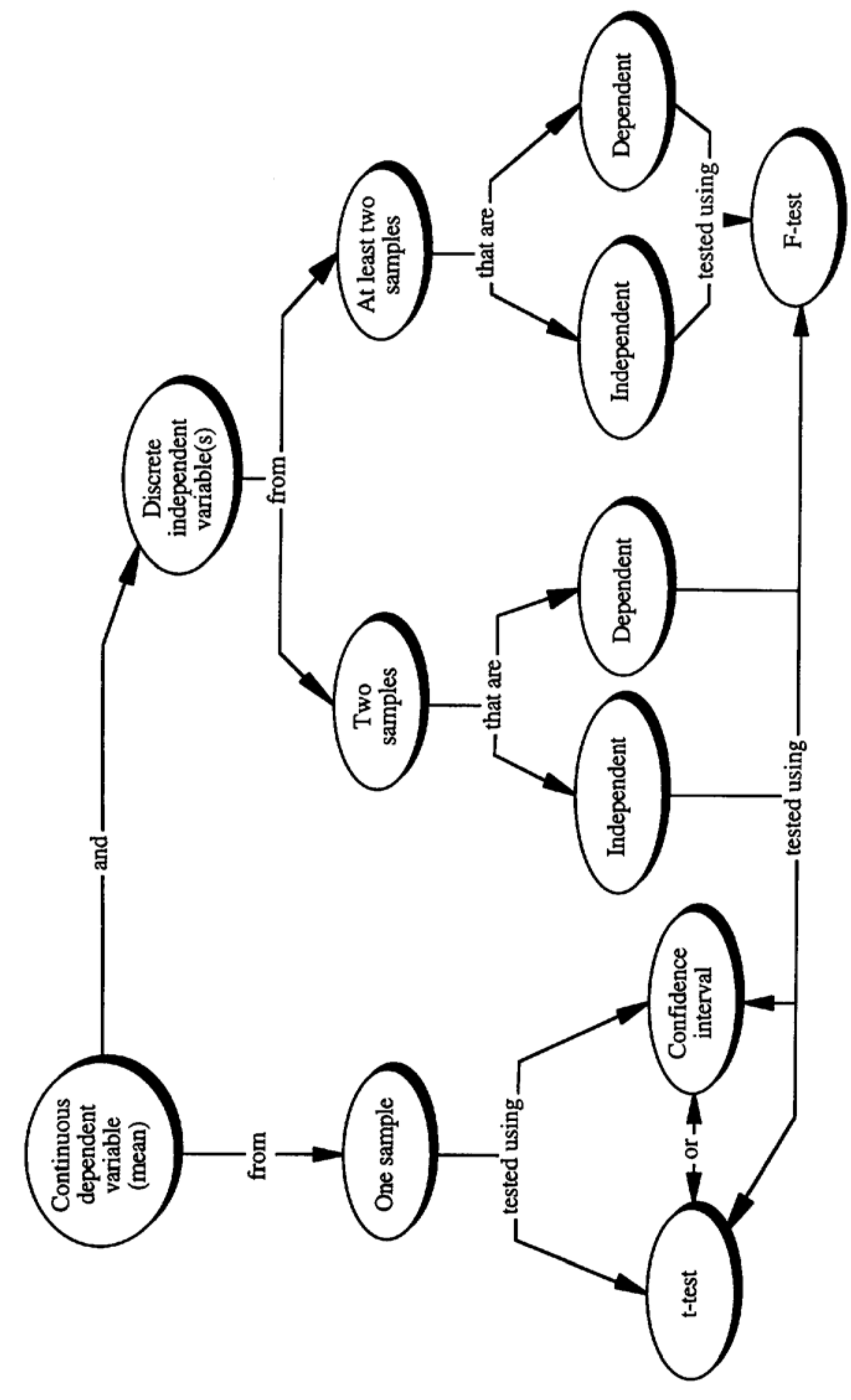

Figure 2. Concept map representing inferential testing of means

Multiple-choice items

Many multiple-choice items are written to assess the quantity of facts students can recall. The following item, for example, assesses recall of a symbol.

The sample standard deviation is symbolized by

a. σ
b. σ^2
c. s
d. s^2
e. SS

The distractors in the above item contain symbols that are correct for other measures of variability. Students will have seen all of these symbols previously in class and in their textbook.

It is likely that most students will attempt to answer this item using fact recognition. However, students who understand that Greek letters are associated with population parameters and Roman letters with sample statistics could use that understanding to eliminate distractors "a" and "b." Even on some recall items, students with connected understanding can have an advantage (if they think to use that understanding).

We can write multiple-choice questions that require connected understanding in order to arrive at the correct answer (unless, of course, a student is lucky enough to guess the correct answer). These items include distractors that lead students with incorrectly connected concepts to select a wrong answer. They are difficult and time-consuming to write. In addition, each item can assess only a small area of the students' networks.

For example, the following item assesses aspects of students' connected understanding of standard deviation.

A group of 30 introductory statistics students took a 25-item test. The mean and standard deviation were computed; the standard deviation was 0. You know that:

a. about half of the scores were above the mean.
b. the test was so hard that everyone missed all items.
c. a math error was made in computing the standard deviation.
d. everyone correctly answered the same number of items.
e. the mean, median, and mode from these scores probably differ.

The correct answer is "d." Students who connect standard deviation to the normal curve and z scores may choose distractor "a." Students who confuse the mean and standard deviation may choose option "b." If students have not seen an example or application in which the standard deviation equaled 0, they may decide that this value is impossible to obtain and choose option "c." Students who remember that values for the three measures of central tendency usually have differed in their examples may choose option "e."

We used this item on the first test in our introductory graduate-level applied statistics course. Students were allowed to use extensive notes but not their textbooks. About 75% of our students answered this item correctly. It discriminated well between higher and lower achievers; that is, many more students with higher total test scores answered this item correctly than did students with lower total test scores.

Word problems

Word problems may be the most frequently used format for assessment in K-12 mathematics instructions as well as in small post-secondary mathematics and statistics courses. Typically, the student is presented with a set of raw or summary data and asked to do something with that information. We used the following word problem on the first test given in the course we described above. Students had worked similar problems, but none was identical to the test item.

> Telia scored at the 90th percentile in grade 8 on a standardized test given at the beginning of the school year. The mean grade-equivalent (GE) score of the 8th-grade norm group on this test is 8.0 and the standard deviation is 2.0 GE. What was Telia's GE score?

The solution to this problem involves transforming Telia's percentile score to a z score using a unit normal (z) distribution table, then solving for Telia's GE score using the z score formula. Again, our students had access to notes.

Our students tried to build three common mental models. Students with accurate connected understanding first recognized that this problem involved the connections among percentile scores, the normal curve, and z-scores and proceeded from that recognition. Students without connected understanding used one of two approaches. Some tried to recall both steps as well as the z-score formula (if they were able to progress that far). Others looked for similar problems in their notes and then tried to use the identical approach in solving the test problem. Thus, students with little or no connected understanding can find ways to correctly answer these kinds of word problems. In fact, our students find problems of this type much easier than our connected understanding multiple-choice items.

Word problems can be written that require connected understanding for correct solutions. For example, students could be given all of the information found in a typical word problem *except* the question they are to answer. They could be asked to generate one or more questions that could be answered from this information. A skeleton of such a problem follows.

> From a research study, you have determined that the relationship between math scores and verbal scores on an achievement test is .60 based on a sample of 45 eighth-grade girls and .70 based on a sample of 35 eighth-grade boys. Generate three questions (in word and hypothesis forms) that could be answered using this information.

We have used problems such as this one as part of our classroom instruction but have not yet included them in formal instructional assessment or in our research.

USING CONCEPT MAPS TO ASSESS CONNECTED UNDERSTANDING

Concept maps are valuable for both summative and formative assessment of connected understanding. As instructors, we want to know how well our students understand statistics at the end of our coverage of topics and eventually at the end of our statistics course; in addition, most of us have to assign grades that reflect our judgment of their understanding. Concept maps can provide this kind of summative assessment. As importantly, though, we need to know how students' cognitive networks develop and change throughout their exposure to statistics. Concept maps can provide this formative information also. Using them formatively allows us to alter instruction to benefit our current group of students; using maps summatively benefits our current class in consolidating their learning and our future classes by indicating areas that need more instructional emphasis.

Two approaches have been used to assess connected understanding with concept maps. In the first, students draw (or create) their own maps. In the second, students complete an incomplete map. Other forms of visual representation (e.g., flowcharts) can be used with either of these approaches, although we have not explored their use in this chapter. See Jonassen et al. (1993) for an extensive presentation and discussion of the use of various kinds of maps for assessment.

Map creation

Novak and Gowin (1984) developed concept maps for use in research and evaluation in science education. Since that time, concept maps that have been drawn directly by learners have been used as measures of connected understanding. Students arrange important concepts into a map and connect them with links they label. Novak and others have developed various quantitative scoring systems for these maps. Points are awarded based on map characteristics, including number of correct propositions, levels of hierarchy, cross links, and examples.

In an introductory statistics course, for example, students could be asked to create a concept map that includes the most important concepts and their connections from a unit they have completed. Students can be given the concepts or asked to generate them. They can complete this task individually or in groups. The maps can be drawn by hand, on computers, or by arranging note cards. Other variations exist; see Ruiz-Primo and Shavelson (1996) for a summary of concept map formats and scoring variations as used in science education.

There is growing research evidence, based on expectations from theory, that the creation of concept maps can be a valid measure of connected understanding for many students (e.g., Horton, McConney, Gallo, Woods, Senn, and Hamelin, 1993; Novak and Musonda, 1991). First, research findings indicate that the maps of experts are larger, more complex, more connected (especially across neighborhoods), and qualitatively more sophisticated than those of novices. There also is evidence that the structure of students' concept maps become more like those of experts (often their instructors) with increasing educational exposure.

Map completion

Almost any map can be used as the basis for a completion, or fill-in, assessment format. The general process involves first constructing a master map. Keeping that map structure intact, some or all of the concept and/or relationship words are omitted. Students fill in these blanks either by generating the words to use (we call this format "generate-and-fill-in") or by selecting them from a list which may or may not include distractors (we call this format "select-and-fill-in"). Figures 3 and 4 are examples based on the two master concept maps found in Figures 1 and 2. Figure 3 is a select-and-fill-in map; Figure 4 is a generate-and-fill-in map. Surber (1984) may have been the first to use this approach.

We applied the select-and-fill-in approach to concept maps for our work. For a small pilot project in our graduate-level introductory applied statistics class, we created a global master map that was similar to Figure 3 but much more complex. It included 40 nodes; we removed 17.

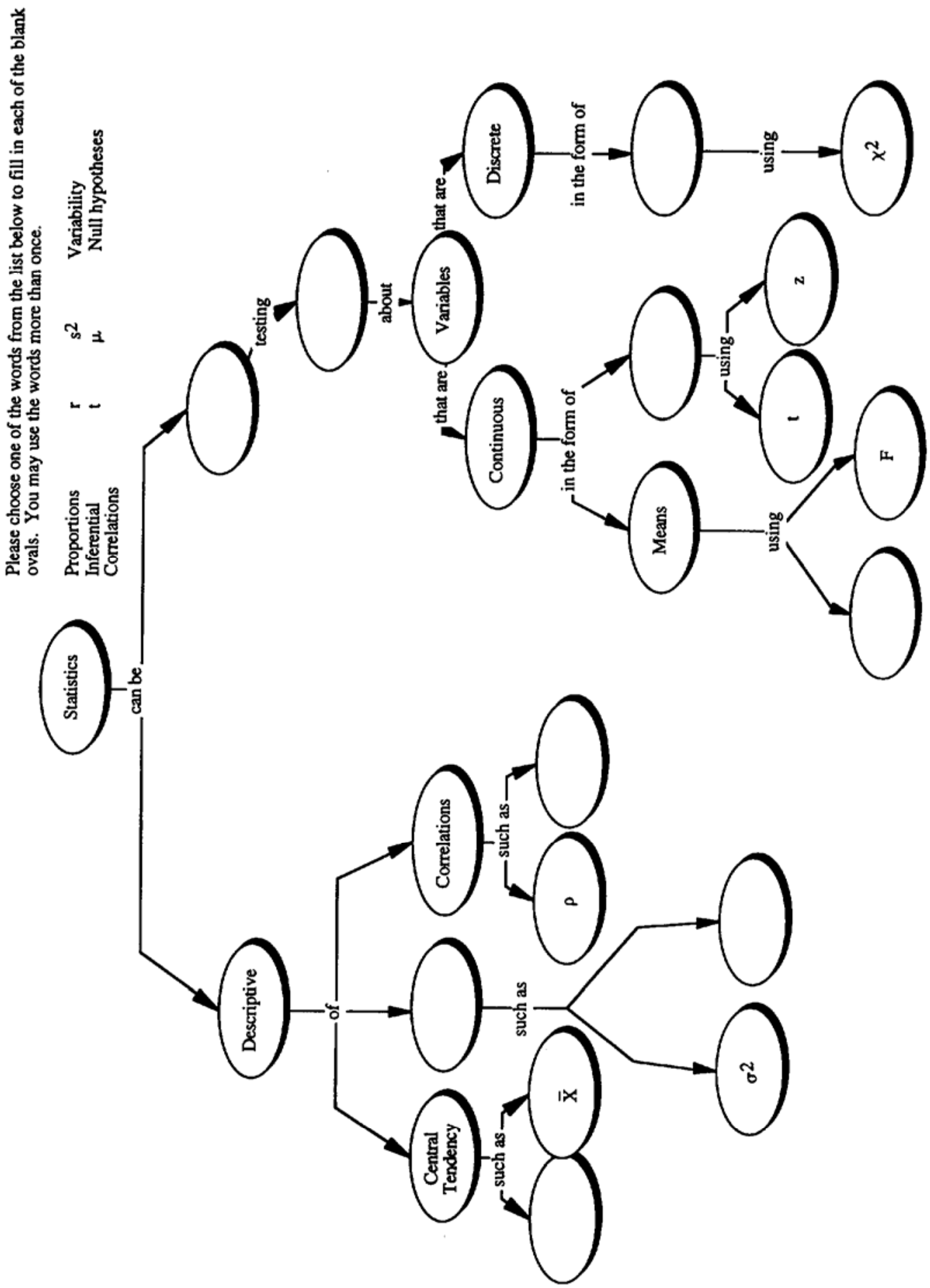

Figure 3. Select-and-fill-in assessment based on Figure 1

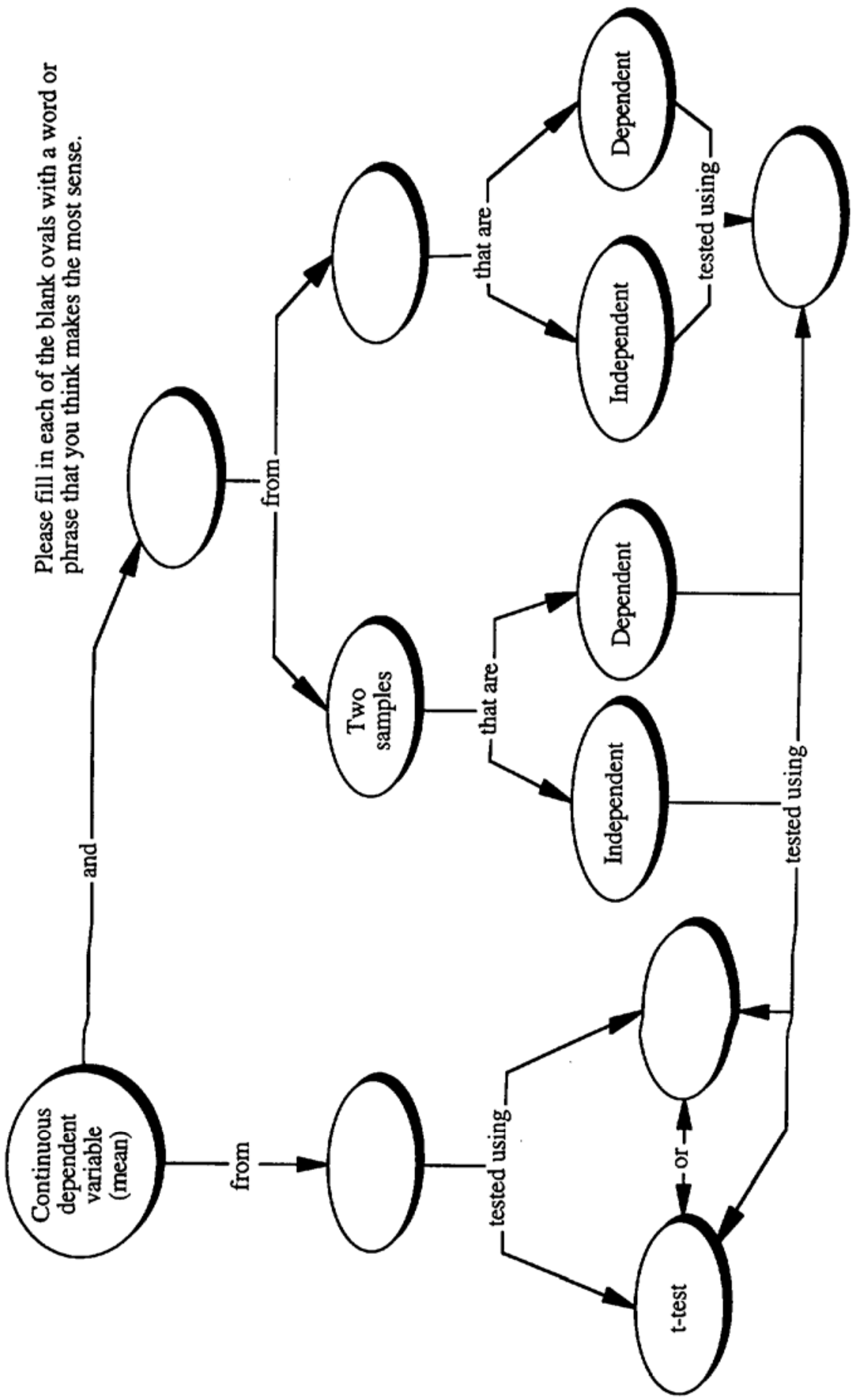

Figure 4. Generate-and-fill-in map based on Figure 2

Students were asked to complete the map at the beginning (pre) and at the end (post) of our course. They selected their answers from a list found on the map; they could reuse the words as often as needed. Map scores were not used as part of our grading scheme, and students voluntarily participated in the project.

We scored each response as correct or incorrect (although other scoring schemes are possible). We examined changes in scores from the pre to the post administration, for the class as a whole as well as for individuals.

Some of these students had taken an introductory statistics course when they were undergraduates, although in some cases that was two decades ago. Even so, at the beginning of the course, students correctly filled in, on average, 25% of the nodes. As a group, these students entered the course with some understanding (some correct, some incorrect), rather than knowing nothing.

At the end of the course, on average, students correctly filled in 73% of the nodes, a large improvement from the pretest average. In fact, most students' post scores were two or three times greater than their pre scores. We also found that these post map scores were strongly related to total points in the course; total course points were based on homework, quizzes, and tests, all of which also were designed to measure interrelated conceptual understanding.

At the end of the course, it was clear (and expected) that most students did not understand all of the important conceptual connections. As instructors, we will attend more closely to these in future classes. If we had administered these maps as a formative assessment during (rather than at the end of) the course, we could have addressed these structural misconceptions during the remainder of the course.

These patterns suggest that select-and-fill-in maps measure understanding of statistical knowledge. Students exhibited a very large gain in connected understanding of introductory statistics concepts across the course, and their fill-in scores were related to achievement in the course. Naveh-Benjamin, Lin, and McKeachie (1995) obtained similar patterns of results using a different kind of fill-in map structure.

SUMMARY AND IMPLICATIONS

From research and instructional experience, we know that many, if not most, people in the U.S. cannot engage in effective statistical reasoning, even after instruction in statistics. One viable reason is that students try to learn statistics (and mathematics) by memorizing formulas. When faced with the conventional kinds of word problems common in statistics classes (which rarely, if ever, look like problems found in everyday life or in research), they attempt to recall a formula, substitute into it, and calculate "the" correct answer. It also appears that many teachers instruct using this approach. This approach to learning and instruction posits that understanding in statistics consists of a collection of relatively isolated facts and formulas; this assumption is clearly untrue.

According to theories about schemas and mental models, as well as work on national standards and goals for mathematics, understanding requires a cognitive network of connected knowledge. These networks can be represented visually as a map of connected concepts. We use the concept map format: a map of labeled concepts connected by labeled links. This model leads easily to assessment formats based on concept maps.

There are two main approaches to the use of concept maps in assessment. The most common approach requires students to draw their own maps; the second, an approach receiving increasing attention, asks students to complete an incomplete map.

Maps can be created in a variety of ways. Students can either generate the concepts they include in their maps or use concepts that are given to them. The concepts can be put on movable pieces to encourage students to try out various structures before settling on one or more. This work can be done individually or in groups. Students generate valuable discussion about the concepts and their connections when they present and/or display their maps. We believe that this flexibility makes them most valuable as a learning tool and as a dynamic (formative) assessment tool, one that qualitatively informs the instructor about areas of understanding and misunderstanding that can then be addressed.

Asking students to generate concept maps has limitations as a summative approach to assessment. First, there is no simple and accepted scoring system for quantifying the quality of concept maps (Ruiz-Primo & Shavelson, 1996). Second, students must learn to draw concept maps. Based on our experience, this process can become tedious and frustrating since it often takes several revisions before arriving at an adequate map. Third, it takes time to generate adequate concept maps. Fourth, the process of drawing concept maps creates a high level of cognitive demand, both visual-spatial and verbal. Fifth, maps are unique to each individual since each person's cognitive understanding is unique (although, of course, there are similarities among experts' maps). Sixth, some students (and instructors) do not like to draw concept maps.

We are experimenting with map completion as both a formative and a summative assessment format. These map completion formats have several advantages. First, an easy score to obtain, and one that is understandable to others because it is familiar, is the number (or percent) of correct responses. Second, fill-in formats can be designed to be easy or difficult. Harder assessments are created by constructing more complex master maps, by omitting more words, by omitting words that are close to each other in the map, and/or by requiring students to generate their responses rather than select them. Third, fill-in formats can be administered relatively quickly and to large groups of students. Fourth, they do not require a computer for administration or scoring, although computers can be used for these purposes. Fifth, fill-in assessments may well involve a learning component since students see the map structure as they complete the task. Sixth, we have evidence that students of almost all ages, with 5 to 10 minutes of training, can and will try these fill-in structures. Seventh, fill-in maps look like they measure connected understanding; this "face" validity is an important consideration in the use of assessments in the classroom, although not necessarily a concern in research.

Another set of emerging techniques for measuring connected understanding that we have not tried to use in our statistics education work is characterized by a two-stage indirect approach. These techniques have been used primarily by researchers, often psychologists, interested in studying cognitive structures; they rarely have been used in instruction. As one example, Johnson, Goldsmith, and Teague (1995) have run a series of research studies in which students rate the degree of connection or "relatedness" between pairs of concepts. Used in statistics, for example, students could be asked to rate the relatedness between pairs of concepts selected from those found in Figures 1 and 2. Using these ratings to represent the "cognitive distance" between each pair of concepts, students' cognitive networks can be represented visually (in the form of a map) and numerically (as scores indicating the agreement between experts' and students' ratings). Both representations are easiest to obtain using computer software programs. These techniques have a number of limitations for use in the classroom. In our view, their major disadvantages are in their dependence upon technology and the problem of convincing students that the tasks are worthwhile. At this point in their development, these kinds of techniques make better research tools than classroom assessment measures.

Before selecting a format to assess connected understanding, you need to consider your model of learning and thinking and determine the purpose for your assessment. Teachers and researchers will want to use a variety of kinds of measurement for different purposes. The types of connected understanding assessment formats we have described in this chapter complement, but do not replace, other kinds of achievement assessments. Many people believe that assessment drives instruction. If so, we hope that the use of these techniques will encourage statistics teachers and researchers to consider the connections that are inherent among statistical concepts and how to encourage connected understanding in students.