

An Annotated Dissertation Proposal Using Quantitative Methods

The choice of topic in the following dissertation proposal, a study of the effectiveness of concept mapping in improving problem solving, is typical of topic choice in doctoral dissertations. Undoubtedly, Kathy Beissner had an “itch to scratch.” Since the improvement of problem solving is central to the work she does as a trainer of physical therapists, why not tackle it in her doctoral dissertation? One must give her credit for undertaking a difficult problem central to her work. Further, where researchers so often work on abstract problems primarily of interest to other researchers, Kathy’s problem is for those on the therapist-training front line.

Now comes the “but,” however. An individual’s problem-solving skill is developed over a lifetime; in the case of Kathy’s students, it was over the past eighteen to nineteen years. Her intervention, by the constraints on her own time and resources, must be comparatively small. Eisner, in Bracey (1994), noted that the length of an experimental intervention in 1981’s *American Educational Research Journal* averaged only seventy-two minutes. So Kathy’s intervention will probably not be atypical. But, as Bracey (1994) notes, “this is a minuscule amount of time when placed against the enormous blocks of time represented by a school year.” And, one might add, it’s even smaller in contrast to such a long-term developmental process as this! From just the title we don’t yet know the length or the exact nature of the intervention, but Kathy has already set the problem in a context that, while a common one for graduate students, presents difficulties in designing a study sufficiently sensitive to show any effect at all, let alone one that would have any practical significance.

Kathy’s choice is the dilemma both graduate student and faculty face: how does one define a topic with enough “bite” to be satisfying and interesting, to be more than an exercise by having practical ramifications, keep within the scope of the student’s skills and resources, and avoid areas where even top researchers have not yet found a satisfactory approach? Kathy has chosen to err on the side of possible practical significance—assuming that even a small intervention effect could later be developed into something worthwhile. Her faculty

chair and committee, in approving this proposal, apparently decided they could live with this choice as well.

Each doctoral student must make that topic choice: finding a problem within their competencies with a reasonable and feasible approach, yet significant enough they are not just content to work on it, but sufficiently committed to follow it through to the end. Then they must convince their committee of this choice as well.

A STUDY OF THE EFFECTIVENESS OF CONCEPT MAPPING IN IMPROVING PROBLEM SOLVING

A Dissertation Proposal

by Katherine L. Beissner

The introduction in paragraphs 1–3 follows the common pattern of spiraling through the proposal's topic, in this instance, three times: in its most abbreviated form in the title, in a general way in the first paragraph, and in more detail in the following two. A very useful format! It is used often in news articles because it plunges the reader into the topic immediately, but then supplies the details that create a more complete understanding.

Since opening sentences set the tone for the reader, it is important to frame them so they draw the reader in from the start. For example, compare this alternate beginning with the original: "Selecting the correct solution to the patient's health problem is the care professional's key skill; how do you significantly improve it? This project will investigate. . ." Is that better at putting the significance up front and making the reader want to read further?

1. Problem solving, the identification and resolution of patients' problems, is a primary responsibility of health care professionals. As an essential element of practice, attainment of problem solving skill is an important goal in the educational preparation of health professionals. This project will investigate the effectiveness of a study strategy in improving students' ability to solve problems in a physical therapy content area.

2. Prior research on improving health professional students' ability to problem solve has focused on instructional or curricular interventions. One method involves a restructuring of the professional curriculum into a problem-based format, using case studies to present necessary information within the patient-based context in which it will be used (c.f. Shahabudin, 1987; Schmidt, Dauphinee, and Patel, 1987; Norman, 1988). Another strategy is to teach the problem solving process in a step-wise manner (May and Newman, 1980; Olsen, 1983; Jenkins, 1985), providing students with exercises for practice of each of the steps in the problem solving method.

3. Each of these methods focus on the instructor or instructional materials to achieve improvements in learners' performance. In contrast, this project will focus on a study strategy that can be used by learners independently. Study

strategies, as the term is used here, are those techniques used by students to process information from text or other learning materials in an attempt to increase their ability to recall and use the information at a later date. As a learner-based intervention, study strategies can be used with any type of instruction.

Problem Solving

Paragraphs 4–8 are as neat an example of a “funnel” approach to describing the research problem as one could ask for. Note how Kathy starts with a broad, general definition of problem solving, applies it to the narrower medical arena, delineates within that arena what she calls “externally controlled” problem solving from what she intends to study, narrows further to that aspect of the clinical process that seems critical but amenable to change, and finally relates that aspect, the organization of knowledge, to the concept mapping strategy she intends to use to improve it. Like the format used in the introduction, this too is a useful one worth copying.

4. In general terms a problem is “a situation in which an individual is called upon to perform a task not previously encountered and for which externally provided instructions do not specify completely the mode of solution” (Resnick and Glaser, 1976, p209). This definition can be applied to the health care setting by considering a scenario in which a health care professional is confronted with a patient who has a combination of signs and symptoms that are unfamiliar to the practitioner. The patient presents the health professional with a problem, or a “task not previously encountered.” The second aspect of the definition of a problem “for which externally provided instructions do not specify completely the mode of solution” is also met in this scenario. While a general series of questions may be asked of the patient, and a routine set of evaluative procedures performed, these questions and tests do not guarantee that the correct diagnosis is made, or that appropriate treatments will be selected. The health care professional must judge the significance of the information gathered, and integrate the information in order to derive a diagnosis and treatment plan. Problem solving in the health professions is the identification of patient problems based upon the patient’s complaints, history and objective evaluation, and the development of treatment plans to resolve those problems.

5. Other applications of “problem solving” in medicine do not fit this general definition of a problem. For example, when expert systems or flow charts are used to specify which tests to perform and to identify the significance of the test results, providing a diagnosis and treatment plan, “externally provided instructions” do specify the solution to the patient’s problem (c.f. Essex, 1978). When this type of system is used, for the purposes of this research, the process used is not problem solving, because the solution to the problem is generated by something outside the practitioner.

6. Problem solving in the health professions has received much attention in recent literature. The clinical reasoning process, the thought processes used when diagnosing patients and designing treatment programs, has been described in

terms of differences in the ways that experts and novices approach problems. Experienced clinicians appear to have mastered the diagnosis and treatment planning process by learning how to group, or integrate data from patient evaluations easily, and by selecting key questions throughout the interview process (Cutler, 1979). Inexperienced clinicians, on the other hand, tend to collect a great deal of data regarding the patient, and later try to piece together information to fit a particular diagnosis.

7. The need to integrate information is a key component of clinical problem solving. Therefore, study strategies that emphasize integration of content seem particularly appropriate for facilitating clinical problem solving. Spatial learning strategies are a method of representing content structure through the use of two-dimensional diagrams. When studying, these strategies can be used to depict the relationships between ideas. Learning concept interrelationships is speculated to improve learners' ability to draw inferences based upon those concepts (Mayer, 1988). This drawing of inferences is a skill necessary in problem solving. Additionally, spatial learning strategies allow one to represent the relationships between new information and prior knowledge. To do this, concepts from prior knowledge are included on the spatial representation as it is drawn. The linking of new information to prior knowledge is speculated to assist in the transfer of learning to new situations (Mayer, 1988), another skill requisite in problem solving.

8. Concept mapping is one type of spatial learning strategy that may be useful in improving problem solving. Concept maps consist of a hierarchical arrangement of concepts, with labeled lines depicting relationships between the concepts (Novak and Gowin, 1984). This project will examine the effectiveness of concept mapping in improving students' ability to select appropriate treatment plans.

Avoid abrupt transitions; make the write-up flow by adding sentences like this to smooth it: "The next section translates this goal into the operational terms through which the study will be carried out."

Study Variables

Dependent Variables

9. The dependent variable in this study will be scores on a test of problem solving ability in a physical therapy subject area. This researcher-generated test will consist of items requiring subjects to identify appropriate physical therapy treatment techniques for given patient problems and to describe their rationale for treatment selection.

In paragraph 9, the student adds the development of a measuring instrument to the work of carrying out the research study itself. Can she do both? The development of measures often merits a study of its own. Readers will want some assurance that the researcher is capable of undertaking this supplemental task and that its feasibility has already been

tested. Sample items in the appendix that included a plan to pretest them would help allay these concerns.

Independent Variables

Clarity of organization is important since it helps the reader keep the parts in mind and shows their relation to one another. In this section, treatment is discussed at both the beginning and the end. Paragraphs 10–14 would be better organized if they were arranged into two subsections: (1) a subsection on treatment that includes paragraph 10 describing the primary independent variable, and paragraph 14, a measure of the treatment, and (2) a subsection on control variables that includes paragraphs 11–13. Further, the latter might be expanded a bit to discuss the variety of alternative explanations not eliminated by the control group and then indicate the most important ones chosen to be controlled in this study. Such a discussion section might well have resulted in the discovery of the tightened control suggested in the comment on paragraph 26 below.

Remember, the proposal indicates one's ability to handle the problem. A rather complete exploration here of possible confounding factors that might prevent one from attributing the effect to the treatment is an important way of doing this.

Paragraph 10. Is group assignment the primary independent variable? No, it is training in concept mapping strategy. Group assignment is part of the control strategy to eliminate alternative explanations, and paragraphs 11–13 are more of this strategy.

10. The primary independent variable will be group assignment. Fifty percent of the subjects will be trained in the concept mapping strategy prior to introduction of a text passage and testing. The other fifty percent will receive the text passage and testing, but no training in the concept mapping technique until after completion of the study.

11. An inability to recall requisite facts about a patient condition or treatment technique can lead to an inability to select appropriate treatments. Therefore, a second independent variable to be included in this study is a test of factual knowledge on the basis for physical therapy treatment selection.

12. While the use of concept mapping is speculated to improve problem solving ability, other factors that may influence ability to problem solve are also considered in this study. First, a subject's critical thinking ability may impact his/her ability to problem solve. Previously acquired abilities to draw inferences and to make judgments of evidence may influence a subject's ability to critically read draw conclusions from text. The Watson-Glaser Critical Thinking Appraisal (Watson and Glaser, 1980) will be used to assess critical thinking ability.

Paragraph 12. Good, a well-known test of critical thinking is being used. Use well-researched instruments, where they are appropriate; they make results more understandable and give readers more confidence in the results. Where new or obscure instruments are a better fit to your variables, give as much description of the measures and their psychometric characteristics as possible so as to give readers confidence in them.

13. Another factor that may influence performance on a problem solving test is the manner in which subjects normally process text. The value of the concept mapping technique is that it requires the map-maker to identify relationships between concepts within text. As used in this study, concept maps will also require that the new content be related to prior knowledge. Subjects who normally study in such a way as to identify relationships in text content can be described as "deep processors" (Schmeck, 1983), because they actively manipulate the content to inner meaning beyond that represented in the text. Those who, in their usual study methods relate new information to prior knowledge can be referred to as "elaborative processors" (Schmeck, 1983). As noted previously, relating new information to prior knowledge is expected to improve ability to transfer learning to new situations, and identifying relationships between concepts helps in drawing inferences (Mayer, 1988). Since both of these skills are required for effective problem solving, a measure designed to detect whether subjects routinely process in this manner is included in the study.

Paragraph 13. Here are two covert behaviors, elaborative and deep processing, which are going to be difficult to tap. How will she determine an individual's routine use of these behaviors? The last sentence is vague on this point. A bit more detail would help, such as: "The Inventory of Learning Processes (Schmeck, 1983) has scales that will be used for this purpose." Give your readers enough detail to keep them positive about your project; avoid vague phrases that may raise questions in their minds.

14. Concept maps generated by learners can vary greatly in their detail and degree of integration of content. Concept maps can be scored according to a standardized method, resulting in a score that indicates the complexity and integration of the concept map (Novak and Gowin, 1984). The concept map scores of the experimental group will be an additional independent variable.

Paragraph 14. Her reference to Novak and Gowin's method lacks indications of its validity, the interrater reliability of scoring, and any evidence of its acceptability by the research community—e.g., use in other studies, studies by other researchers, etc. Additional details are needed somewhere such as the literature review. Then include a sentence like: "As noted in the literature review, this is the most widely used and carefully detailed method of scoring concept maps."

Hypotheses

15. In light of the previous discussion, the following hypotheses are offered:

1. Subjects who use concept mapping to study a text passage on physical therapy treatment procedures will identify appropriate treatment plans for more patient simulations than subjects who do not use concept maps to study the test. H1: concept Map > Control on Problem Solving

2. If concept maps are valid representations of the study processes used when generating maps, experimental subjects (Exp) who develop more integrated and elaborated concept maps (High CM) should outperform subjects with less integrated/elaborated maps (Low CM) on tests of problem solving. H2: Exp: High CM > Low CM on Problem Solving
3. Control subjects who deeply process new information (High DP) should outperform control subjects who do not deeply process new content (Low DP) on tests of problem solving. H3: Control: High DP, Low DP on Problem Solving
4. Control subjects who elaboratively process new information (High EP) should outperform control subjects who do not elaboratively process new information (Low EP) on tests of problem solving. H4: Control: High EP, Low EP on Problem Solving
5. Subjects with high critical thinking ability (High CTA) should outperform subjects with low critical thinking ability (Low CTA) on tests of problem solving. Hs: High CTA, Low CTA on Problem Solving

Paragraph 15. Note how her hypotheses follow nicely, just as they should, from the discussion of problem solving and description of the variables. No surprises! Great, there shouldn't be!

Limitations

A limitations section, paragraphs 16–18, is often the place where the student seeks to forestall the readers' questions about the generalizability of the study. It often contains all kinds of disclaimers, many of which are so far-fetched as to be unnecessary. While not far-fetched, paragraph 17's disclaimer is probably unnecessary, since in many ways this study could be considered research on the improvement of problem solving in general; it just happens to use health professions material and students.

Paragraph 16, however, is an important limitation because only part of the clinical problem-solving process will be involved. Restrict a "limitations" section to similarly important caveats.

16. Since a major factor in problem solving is prior experience, and providing students with identical experiences is not feasible, it is necessary to select subjects with little to no experience in physical therapy treatment. With this lack of experience there is a concomitant problem in that the knowledge base upon which the subjects can problem solve is also limited. Therefore, only a portion of the problem solving process will be tested in this study. The full process consists of identifying the cause of the patient's problem, and then selecting an appropriate treatment. The subjects in this study do not have the background required to diagnose patients' problems. They can, however, be provided with the information required to select basic treatment techniques. Therefore, this study will be limited to an as-

assessment of the effects of learning strategies on the treatment-planning component of problem solving.

17. While elements of the clinical problem solving process are utilized in all health professions, for practical purposes students from only one profession will be used in this study. While there is no evidence to suggest substantial differences between students in the different health professions, if such differences between professions do exist the inclusion of only physical therapy students in this study may limit its generality.

In contrast to the abrupt transition earlier, note the nice transitional paragraph, number 18, that helps us anticipate what is coming and how it is organized. Transitional material helps keep readers oriented and integrate their thoughts about the proposal.

18. The next section of this proposal deals with the literature related to the proposed study. First, problem solving in the health professions is examined, and the research on instructional methods to facilitate problem solving is discussed. Since the cognitive processes that are activated when using study strategies may have an effect on the learning outcomes, the literature on cognitive processing is given an overview, and the cognitive processes required in problem solving are identified. This information is then related to the literature on learning strategies, focusing on strategies that represent relationships between ideas in graphic formats.

Review of Related Literature

19. Clinical problem solving requires "the ability to gather data from the patient by history and physical examination, integrate this information into a diagnostic formulation, select appropriate investigations to confirm the diagnosis, and institute efficacious management" (Norman and Feightner, 1981, p. 26). While the outcome of the problem solving process, the treatment plan, is the same for experts and novices, the thought processes that occur en route to a treatment plan may differ. When attempting to diagnose a patient's problem, experts often rely on prior experience with similar patients to generate hypotheses about the source of the patient's problem. These hypotheses, which are generated early in the evaluation process, are then used to guide the remainder of the evaluation, as information is gathered to confirm or refute the hypotheses (Barrows and Feltovich, 1987). Likewise, in the formulation of treatment plans, experts rely on past experience to determine what treatment approaches to use. This is in contrast to the novice clinician who has little prior experience on which to base hypotheses or select treatments. For the novice, each new patient is seen as a completely new "problem" that must be investigated.

20. Expert clinicians can be seen as using something other than a traditional problem solving process (Norman, 1988) because each new patient can be seen as similar to some prior patients. In the diagnosis and treatment process each patient is not, in Resnick and Glaser's description of a problem, "a task not previously encountered" (1976, p. 209). In contrast, novices do evaluate and treat patients ac-

cording to a problem solving method, since each patient is seen as a new "task." This problem solving process requires that information regarding the patient be gathered and integrated to formulate a diagnosis, or a list of symptoms that can be treated. Then the clinician must recall the various types of treatments that are available for treating the patient's problem, and integrate the information regarding each treatment's effectiveness with information from the patient evaluation to select an appropriate treatment regime.

Methods of Promoting Clinical Problem Solving

21. If the purpose of professional education is to train individuals to be problem solvers in their respective fields (Cyert, 1980), the educational preparation of health professionals must address this skill. Attempts to improve health professional students' clinical problem solving ability have taken several forms.

Note that in paragraph 21, Kathy goes on to another topic without tying the relevance of the expert-novice contrast discussed in paragraphs 19 and 20 to her study. Apparently, experts and novices use different paths to problem solving since the latter doesn't yet have categories for diagnosis. So what? Does that invalidate the use of her students? What are we to conclude?

The point of the literature review is not to cover everything that was said on the topic but to demonstrate how well you have chosen references that have some bearing on your study. It should display your competence in mastering this subject matter and original thinking in revealing that relevance. Until they are related to the study, paragraphs 20 and 21 don't do this.

22. [The next eighteen pages of literature review up to its summary, paragraph 23, were omitted.]

Summary

23. In summary, three distinct arguments can be made for using spatial learning strategies to promote problem solving ability. The first is made on the basis of the transfer-appropriate processing hypothesis (Bransford, Franks, Morris and Stein, 1979; Morris et al, 1977). According to this hypothesis, efficient studying requires that the initial processing of material be compatible with the ultimate testing conditions. Construction of spatial representations requires selection of important material; in problem solving one must be able to attend to relevant cues, disregarding extraneous content. Clinical problem solving requires that relationships between problems and symptoms be identified, similar to the building of internal connections and labeling of relationships used in spatial strategies. When relationships between concepts are not explicitly stated within the text of the material to be learned the studier must infer the nature of the relationship. Inference is essential in problem solving, as one must determine the cause of the patient's symptoms. It is

argued here that the cognitive strategies required for construction of a spatial representation of text material are consistent with the cognitive processes required by problem solving, indicating that these strategies are appropriate study tools for improving problem solving ability.

24. Second, according to Mayer's "qualitative" view of learning, inference is enhanced by study strategies which require building internal connections between concepts in the to be learned material. As noted above, when working with ill-structured problems such as those encountered in the health professions, inference is an essential element of the problem solving process. Likewise, building external connections between the to be learned material and prior knowledge enhances learning transfer, resulting in improved ability to solve novel problems (Mayer, 1984, 1988). Transfer of learning to new situations is an essential element of problem solving. Strategies that require generation of both internal and external connections have been shown to assist in problem solving (Bromage and Mayer, 1981; Mayer and Cook, 1984). Both internal and external connections are depicted on concept maps.

25. A final argument for the use of spatial learning strategies in general, and concept mapping specifically, for the improvement of problem solving skills, is that such strategies assist in representing the structure of the subject area. Knowledge of content structure is speculated to assist in transfer of learning and problem solving ability (Bruner, 1960). Mental models (Mayer, Dyck and Cook, 1984), the mental representations of relationships between elements in a content area, provide the background structure on which inferences can be drawn and problems can be "run" (de Kleer and Brown, 1983). Wittrock (1988) contends that the purpose of the mind is to create mental representations or models of the world, and the mental representations or models are used to solve problems. Spatial learning strategies, then, are seen as ways to acquire knowledge and build these representations.

Paragraphs 23–25. Ah, here we are told the relevance to her study of the topics in the literature review. She very nicely justifies her treatment for improving clinical problem solving! But connections of previous research to her study should be made throughout the review. Summaries should summarize—briefly restate and organize previous arguments, not surprise us with new material.

Methods

The methods section, paragraphs 26–49, is the heart of the proposal; what has gone before is further operationalized. It should cover the six facets of experimental design: participants, situation (participants and situation, as they often are, are linked in this instance), focus of action (treatment and independent and dependent variables), records (observations and measures), comparison and contrast (basis of sensing attributes and changes), and procedure. The proposal structure roughly follows the ordering of these facets.

Note that while "subjects" was appropriate in the day when this proposal was completed, today her advisors would suggest that she use words like participants or inform-

ants to indicate a more equal status with the investigator and, in the latter case, their importance in providing the grist for the mill.

Subjects

26. The study population consists of first and second year college students intending to major in a health profession. Sixty volunteers from the freshman physical therapy class at Ithaca College will serve as the study sample. If insufficient numbers of freshman students volunteer to participate in the study, students from the sophomore physical therapy class will be recruited for the study. Subjects will be told that participation in the study is strictly voluntary. They will be told that by participating in this study they will learn a study technique that may assist them in preparing for their physical therapy courses. Subjects will be randomly assigned to treatment groups using a table of random numbers.

Paragraph 26. Good! Using volunteers for the control as well as the experimental group and random assignment of individuals to groups control for a lot of potentially contaminating variables. Usually, one would also expect to see that the treatment will be randomly assigned to one of the groups, perhaps by flip of a coin. That ensures the researcher won't assign the treatment to the better group to give them a head start (but see footnote in next paragraph). As you will see in paragraph 40, however, she is going to test for factual knowledge of the material on physical therapy treatment that her students are to master as well as their critical-thinking ability. What if, by chance, the experimental group masters the subject matter considerably better than the control group? Or already are better critical thinkers? She would have great difficulty showing her experimental treatment worked.

On average, random assignment will make the groups equivalent. But to avoid the rare case where it fails to do so, she should have sorted the total group into three or four subgroups on the basis of their scores on factual knowledge of the experimental material. Then she should have sorted each of those subgroups on the basis of their critical-thinking scores. With three subgroups on each test, this would yield nine blocks of students. She would then randomly assign students within each block to experimental and control groups. That would have eliminated differential mastery of the reading material as well as differences in initial problem-solving ability as alternative explanations of apparent treatment effectiveness.¹

1. As the dissertation data turned out, random assignment did sufficiently control for these variables—indeed, more often than not it does. But should the experimental group have started out markedly superior, we do not have satisfactory means for correcting for this head start. Rather than assignment of treatment to groups by a coin flip, a conservative approach assigns it to the lower scoring one. The treatment effect is assumed strong enough to overcome the handicap. At best, this yields a conservative measure of treatment effect, however, and it may not be possible where, as here, one is controlling for more than one variable (for more, see Cook and Campbell, 1979, pp. 103 ff.).

Materials

27. The content used as the basis for testing will be an approximately 2000 word passage on the physiological effects of thermal agents (heating and cooling treatment techniques), and a description of the different forms of thermal agents. The text passage is written in a style consistent with that used in physical therapy textbooks, with the citations of supportive research studies deleted, and language modified to accommodate subjects' limited experience with medical terminology (see Appendix A).

Paragraph 27. Great, she has it all ready to go and shows it to us in the appendix!

Instruments

Dependent Variable

Problem Solving. 28. A number of methods of assessing health professional students' abilities in problem solving have been developed. The Patient Management Problem (PMP) is a written test used to determine the method the test taker would use to assess and plan treatments using branched simulations. The simulations begin with a brief description of a patient, followed by questions regarding the type of information needed to diagnose the patient. The test takers' goal is to select appropriate test information to diagnose and plan an effective treatment. The information selected during the PMP is provided to the test taker, who then makes judgments regarding additional data required, the patient's diagnosis and/or appropriate treatment procedures. Information not requested is not provided to the test taker, thus simulating actual data collection procedures.

29. Goran, Williamson and Gonnella (1973) examined the relationship between performance on PMPs and actual clinical performance of physicians as documented in medical records. They found that significantly more tests were ordered on the PMP than were actually ordered in the same physicians' clinical practice, thus reflecting questionable validity of the PMP. Construct validity was established by Sedlacek and Nattress (1972) by having expert judges rate a series of decisions in a PMP in terms of their appropriateness. The inter-rater reliabilities ranged from .71 to .85.

30. Newble, Hoare and Baxter (1982) investigated the differences between orally presented PMPs and written PMPs on the premise that written PMPs provided increased cueing to test takers. They found that medical students and physicians with different levels of experience all requested more information when the PMP was presented in the written format vs. the oral (uncued) format. Further reported findings indicated that it appeared that medical students outperformed the more experienced physicians on the PMPs. Since one would predict that the physicians with the most experience should have the highest scores on the PMP, the authors questioned the validity of the PMP. However, the results on differences

between groups of physicians were not statistically analyzed. Wolf (1984) reanalyzed data gathered by Newble, Hoare and Baxter (1982) to find that the PMPs did not discriminate between the various levels of physicians and students. Thus, while the findings did not support the contention that more experienced physicians should outperform less experienced students, the results did not indicate otherwise for this one PMP. A greater number of PMPs tested on a larger study sample may provide a better measure of the construct validity of the PMP (Wolf, 1984).

31. This type of study was conducted by Farrand, Holzemer and Schleutermann (1982), who examined the validity of PMPs by comparing the performance of certified nurse practitioners to nurses with basic registered nurse (RN) training. Their findings supported the hypothesis that the nurse practitioners' diagnostic performance should be superior to the less highly trained RNs.

32. While PMPs provide one method of assessing problem solving, the technique has been criticized because it is unstandardized, and therefore difficult to construct and use. Each test taker receives different information and in a different sequence, according to the information requested. Another technique used to assess problem solving ability which is more standardized is the modified essay question (MEQ).

Paragraphs 28 through 31 set us up to accept the PMP as a dependent variable measure with a nice discussion that includes paragraph 31 indicating that for nurses, a group for whom the data are more relevant to physical therapists than the data for doctors, there is evidence attesting to its validity. But paragraph 32 veers into a series of problems with the PMP and raises the MEQ as an alternative. This leaves the reader wondering, "Why"?

In fairness to Kathy, we do find some of that information at the end of paragraph 33. There, having described the MEQ sufficiently so the reader will have some understanding of her choice, the PMP and MEQ are contrasted. In part, therefore, this is a matter of style. We very strongly advocate answering any question in the reader's mind raised—if possible, even before. Then their reading of the proposal continues with a positive attitude toward the project instead of it being continually tinged with questions.

But in terms of content as well as style, there is still reason for concern. It is not clear how the fact that all students get the same information is critical, especially since "seeking and choosing" information is part of the problem-solving process in the real world. Perhaps it is that "seeking and choosing" is more closely connected to diagnosis (which Kathy has excluded from her study) than to choice of treatment (her focus). If so, the MEQ may be more desirable. But whatever the reason, she should have explained it to her audience.

33. The MEQ has been used to assess medical (Knox, 1989; Irwin and Bamber, 1982) and physical therapy (Stratford and Pierce-Fenn, 1985) students' problem solving ability. Each MEQ consists of a brief written introduction of a patient case history followed by a number of questions based upon the presented information. The questions asked may test factual knowledge, or may test higher levels of cognitive objectives. Following the completion of the presented questions test takers

are presented with additional data regarding the patient, as well as additional questions to which they must respond. This process of data presentation followed by questioning is continued until all desired questions have been responded to. MEQs are typically presented in booklet format as written tests. Test takers are not allowed to read ahead in the test booklet, nor may they return to previously answered questions to make corrections or changes in their responses. MEQs differ from patient management problems (PMP) in that all test takers are provided with the same information regardless of their responses. In the PMP test-takers receive only the information requested. Therefore, with PMPs the test itself is individualized, while the MEQ is more standardized.

Paragraph 33. As readers, we'd like to know what Knox (1989), Irwin and Bamber (1982), and Stratford and Pierce-Fenn (1985) learned from their use of the MEQ, since they aren't cited again. Add a note like, "... gave no analysis of usage" to allay that question and keep us from wondering if the author had simply ignored their findings.

34. MEQs have been used as a major part of the assessment of students in the Newcastle Medical School in Australia (Feletti and Smith, 1986). In this application the MEQ was intended to be focused on problem solving ability, and the application of knowledge rather than a test of recall of prerequisite knowledge per se. Analysis of the actual MEQs given over a three year period, however, indicate that the percentage of items testing problem solving ability declined, with an increase in the percentage of items testing factual recall and data interpretation (Wolf, 1986). Irwin and Bamber (1982) likewise analyzed the content of MEQ examinations given at Queen's University in Belfast. As in the Newcastle study, the number of items testing higher levels of cognitive objectives (analysis, synthesis and evaluation) declined over a two year period, with a concomitant increase in the number of items testing recall and comprehension. Apparently, when MEQs are given over a period of years there is a tendency to increase the emphasis on factual knowledge, and decrease testing of higher level cognitive skills.

Paragraph 34. Once again nice information, but of what relevance to the project? This isn't a study that would extend over time.

35. The reliability of MEQs has varied according to the content area tests. Internal consistencies of MEQs used with physical therapy students ranged from .75 to .92 when adjusted to a 60-item questionnaire.

36. Disadvantages of the MEQ include the effort required to create an acceptable test (Knox, 1989). A difficulty appears when test takers use different approaches to the clinical problem presented than the test constructor, thus finding themselves out of sequence with the information presentation in the MEQ test booklet. Student feedback regarding the MEQs at Newcastle medical school indi-

cate that the questions are sometimes perceived as ambiguous. Students also claimed that they felt a great deal of time pressure, and were uncertain of the degree of detail required to complete each item (Feletti and Smith, 1986). While acknowledging the difficulties with MEQs, Feletti contends that the test is a valuable standardized problem solving assessment tool, if other instruments are used to assess knowledge and data interpretation (Feletti and Smith, 1986).

We're glad her paragraph 36 alerts her audience to problems with the MEQ. But once again, she needs to relate the material to her study. How will these problems be solved? Will she structure the test so students don't get out of sequence? Or doesn't sequence make any difference in this study? How will she ensure that questions are not ambiguous? Will she pretest to set time limits generously?² Will examples indicate the level of detail wanted? It is getting to be an old refrain, but when questions are raised in the reader's mind, answer them or indicate why you can't.

37. In this study, a modified version of the MEQ will be used to assess subjects' ability to select appropriate treatment techniques. The standard MEQ presents test takers with a brief description of a patient, and, following a series of questions, additional information on the same patient is provided. This provides an intense examination of the ability to problem solve on a single patient.

Paragraph 37. When terms like modified version are used, as in paragraph 37, one needs to clearly indicate the nature of the modification. We infer that her modification of the MEQ is that it will test only the treatment aspect of problem solving, but that needs to be made explicit.

38. As noted previously this study will examine only the treatment planning aspect of problem solving. The subjects for this study do not have the background required to make decisions about simulated patients' diagnoses. The items on the investigator-developed instrument will consist of brief patient descriptions, referred to as case studies, similar to the type used for MEQ. The cases used in the study will be realistic examples of patients with musculoskeletal problems. For each case, an appropriate form of treatment will be one of the treatments covered in the text to be used in this study. Each case will be followed by two questions, asking the subject to a) identify one thermal agent treatment that is most appropriate for this patient, and to b) explain why that particular thermal agent is most appropriate for the patient. Sample patient descriptions of the type to be used in this study are provided in Appendix B.

2. She didn't pretest! In the discussion section of her dissertation, where she is explaining why the map scores did not relate to the dependent variable, she noted insufficient time as a possible reason. Pretesting pays!

Paragraph 38. Good use of the appendix again! But it would be even better if she included at least sample questions. Examples not only indicate careful prior thought but also give the dissertation committee a chance to apply their expertise through making concrete suggestions. After all, specialized know-how is one of the criteria for selecting members of a dissertation committee.

39. Validation of this testing instrument will be made with the use of three experts in the area of thermal agents. These experts will review the cases generated and respond to each item. The responses will be compared with those of the test maker for consistency in treatment selection. The experts' responses to the items will be used as the scoring key for the test. In addition the content experts will be asked to determine whether each item on the problem solving test requires the cognitive processes identified as important to problem solving. They will be provided with a brief description of the cognitive processes required in problem solving. For each item on the problem solving test experts will be asked to indicate whether each of the cognitive processes are required to correctly respond to the item. Those items that do not require the cognitive processes identified as essential to problem solving will be modified so as to include these processes.

Paragraph 39. Another instance—anticipate readers' questions and answer immediately! Here one wonders what happens when the "experts" disagree either with one another or with the test maker. If the readers didn't think of that question in this paragraph, it will be raised in their minds by the last sentence of the next.

Independent Variables

Factual Knowledge. 40. Factual knowledge on the reading will be assessed with an investigator-generated 20 item multiple choice test covering the content of the text. Items on this test will require strictly recall of content, with no inference. Validation of this test will again utilize the content experts, who will individually take the test. Items on which the content experts do not agree will not be used in the study.

Paragraph 40. Multiple choice-test construction is not always as easy as it looks. Pretesting of instruments (surveys, tests, observation forms) is essential to knowing whether they work as intended. In this instance, having experts take the test may have validated the correct responses and their relevance to the reading passage, but it did not serve as a pretest with subjects like those she would be using. The test constructed for the dissertation proved too easy; the mean score on the 17 item test was nearly 16 (no explanation was given for the drop from the proposal's 20 to 17 items). Therefore, the test did not discriminate well enough among the students for there to be a valid measure of factual knowledge. Pretesting, of course, would have discovered this problem before the final data were gathered.

Critical Thinking Ability. 41. Students' innate ability to reason critically may affect their performance on problem solving tests. Therefore, critical thinking ability will be measured with the Watson-Glaser Critical Thinking Appraisal (CTA) Form A (1980). This test is comprised of 80 objectively scored items. Five subscales of the appraisal assess inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments, five elements of critical thinking. The test-retest reliability of the CTA is .73, while split-half reliabilities range from .69 to .85 for the five subscales. Construct validity of the CTA has been established by studies which tested traits presumed to be related to critical thinking ability.

42. This instrument has been used in previous studies of critical thinking in nurses (Matthews and Gaul, 1979; Pardue, 1987; Tiessen, 1987) and physical therapists (Slaughter, Brown, Gardner and Perritt, 1989).

Inventory of Learning Processes. 43. The chief hypothesis of this research is that use of study strategies which include building internal connections between concepts within text and relating content to prior knowledge will improve ability to problem solve. Since some subjects may already use study strategies that include building internal and external connections, the Inventory of Learning Processes (Schmeck, 1983) will be used to identify usual study methods. The Inventory of Learning Processes (ILP) is a self-report inventory which consists of 62 true/false items regarding students' study habits. Four subscales are included in the inventory. "Deep processing" consists of 18 items which assess the extent to which students evaluate, compare and contrast, and organize information as they study. "Elaborative Processing" is a 14 item subscale designed to determine the extent to which students relate new material to prior knowledge. The 23 items on the "Methodological Study" scale assess how much students report that they conform to practices recommended in "how to study" manuals. The final scale, "Fact Retention" is composed of 7 items. This scale assesses students' tendency to adopt study habits that encourage rote memorization.

44. The test-retest reliability of the ILP subscales, when tested over a two-week period range from .79 to .88. Internal consistencies range from .58 for the seven item "Fact Retention" scale, to .82 for "Deep Processing."

45. Test construct validity has been established through tests measuring traits expected to correlate with specific learning process subscales. For example, Schmeck and Ribich (1978) found a significant relationship between the Deep Processing scale of the ILP and total score on the Watson-Glaser Critical Thinking Appraisal. Since deep processing involves careful evaluation of data, and comparisons and contrasts between concepts, this relationship was predicted. Other predicted relationships between the ILP and other standardized assessment instruments include a positive relationship between reports of mental imaging and the Elaborative Processing scale, positive relationships between a measure of academic curiosity and the Deep Processing, Elaborative Processing and Study Methods subscales, and a negative relationship between the Deep Processing scale and anxiety measures (Schmeck and Ribich, 1978). The latter finding is con-

sistent with the Schwartz (1975) study which indicated that highly aroused (anxious) subjects tend to organize memory around superficial aspects of words, such as rhymes, while low arousal subjects organize memory more semantically.

Paragraph 45. The fact that the ILP deep-processing scale and the Watson-Glaser have a "significant relationship" is a positive factor for the validity of the ILP. But it raises the question of what new information the ILP is giving. A careful reader can hardly miss raising this question. It should have been addressed, if nothing else, by indicating the size of the correlation and the reliability of each test. Ideally, also include the proportion of reliable variance on the ILP deep-processing scale that is independent of the Watson-Glaser. (Help from a statistician might have been worth seeking here. Assistance of this and many other kinds is freely available for doctoral students at most universities; make use of them.)

Concept Map Scores. 46. Concept maps generated by the experimental group will be scored according to the standard procedure outlined by Novak and Gowin (1984). This procedure awards points for 1) the number of levels included in the concept map hierarchy, 2) the number of appropriate links identified on the map, and 3) the number of interrelationships identified. The latter measure of interrelationships can be taken as a measure of the degree to which the map maker developed internal connections while studying the material. Since the building of external questions is predicted to have an influence on problem solving ability, an additional measure for assessing the degree to which subjects relate the new material to prior knowledge will be added for this study.

Paragraph 46. How will that score be constructed, and what kind of reliability will it have? Note the number of things yet to be worked through in this proposal—the subject matter test, the MEQs with all their problems, and now this. It isn't overwhelming, but it is enough to raise warning flags for the readers. You want to keep the limit of to-be-worked-through items as low as possible, consistent with the kind of study you are doing. An exploratory study or a qualitative one, typically by its nature, has many. An experimental study, such as this one, typically would be expected to have few.

47. Concept maps will be scored by the researcher and a trained assistant. Interrater reliabilities for concept map scores will be determined using Pearson Product Moment correlations.

Paragraph 47. An intraclass correlation or Cohen's Kappa are the usual statistics for checking interrater agreement. (As noted above, if you aren't comfortable with statistics, have the proposal checked by someone who is; they would catch something like this.) The use of the Pearson product moment correlation to determine interrater reliability will show similarity of ranking of individuals, but will not catch the fact that one is an "easier" grader than the other. Because the intraclass does, it is usually preferred. When using other than the preferred method of analysis, always allay questions by telling why.

In this instance, Kathy is interested only in the relation of map scores to the MEQ. The fact that one rater grades higher than the other makes no difference so long as they agree on the rankings—they did, the correlation given in the dissertation was 0.96. Readers preferring to see an intraclass correlation or Kappa might consider this a flaw unless the proposal includes an explanation.

Procedure

48. Sixty freshman and sophomore physical therapy major volunteers will be solicited via a recruitment letter, and then randomly assigned to either the control or experimental group. Each group will then be divided into two subgroups according to subject availability for meetings. All subjects will participate in 3 sessions. In the first session subjects will be oriented to the study in general terms, and provided with an informed consent form. During this session all subjects will complete the Watson-Glaser Critical Thinking Appraisal, and the Inventory of Learning Processes. During the second session the experimental groups will receive training in the concept mapping technique. The training will follow the suggested outline provided by Novak and Gowin (1984) for training college students, with slight modification. The modification will be to include instructions to relate the text material being mapped to prior knowledge, and to reflect these connections on the concept map. The third session for the experimental group will be the study and testing session, in which subjects will be provided with a text passage on some aspect of physical therapy. Subjects will be instructed to study the passage, and to construct a concept map of the content. Subjects will be allowed one hour to study the material. Then, following a brief break all subjects will be asked to complete the post-test.

Paragraph 48. Don't leave material dangling! She'll divide experimental and control groups into subgroups "according to subject availability for meetings"? What happens then? She uses only those available for all three? Or is this to allow her to adjust session scheduling so subjects can attend them all? Probably the latter, but it isn't clear.

One gets so familiar with the material, particularly toward the end of proposal development, that it is easy to forget how ignorant the reader is. This happens also in the last sentence's reference to a posttest—a posttest of what? One can determine it is the subject matter test, but it should be explicit. Have the proposal read by an "ignorant" reader to catch these.

49. The first session for control subjects will involve completion of the same tests as for the experimental subjects. In the control's second session subjects will receive the text passage, and be instructed to study the material in their usual manner. On completion of the session all notes made during the study session will be collected. Following a brief break, control subjects will take the post-test measure. The third session will consist of training in the concept mapping technique as described above. During each session for both the control and experimental groups

the subjects will be advised not to discuss the content of their sessions with other study participants.

Data Analysis

50. Multiple regression analyses using the method of least squares will be the data analysis procedure. The dependent variable, score on the problem solving post-test, will be regressed against the following independent variables: Group Assignment, Factual Knowledge Score, Critical Thinking Appraisal score (CTA), Inventory of Learning Processes sub scores, and Concept Map Scores.

Independent Variables

Group Assignment(Gr)

Factual Knowledge(FK)

Critical Thinking Appraisal(CTA)

Inventory of Learning Processes Scores

Deep Processing(DP)

Elaborative Processing(EP)

Concept Map Scores(CMS)

The resultant regression equation is:

$$y_1 = b_0 + b_1Gr + b_2FK + b_3CTA + b_4DP + b_5EP + b_6CMS$$

where y_1 = score on problem solving test.

Paragraph 50. Kathy's use of multiple regression analysis in place of the usual analysis of variance is excellent. Not only will it give her information as to which variables are statistically significant, but it will also indicate how much new variance is predicted by each variable. As previously, some discussion of the contrast with the typically expected analysis of variance method with the advantages of this one would be in order. So also would more information about how this will be applied.

Such details indicate one's capability in the area of analysis. A brief treatment, as here, makes the reader wonder whether the analysis is going to be a problem for this student. Allay such questions if you can, either by getting statistical help at the proposal level or by learning enough yourself to handle the problem competently.

Statisticians always prefer to be consulted BEFORE the data are gathered. That way they can avoid many problems that are just not solvable with available methods after the data are already in hand.

Appendixes omitted.

Recall that at the outset we raised questions about Kathy tackling a difficult problem, changing long-standing problem-solving patterns using concept mapping. Curious about the outcome? As it turned out, the experimental group did exceed the control by a statistically significant amount.

But the regression equation explained only 8 percent of the variance, so the effect was

not very strong. Further, none of the other hypotheses were supported, including the fact that scores of the maps did not correlate with the score on the MEQ (her hypothesis 6). Thus, she was left in the anomalous position of having a positive result not supported by other explanatory variables. She offered a number of explanations, many relating to a lack of validity for the measures. Alternatively, her finding may have been a chance result (type 1 error). As Kathy lays out very nicely at the end of her dissertation, a number of additional studies (which she carefully describes) are needed.

So, our initial concern about being able to design a sensitive enough study to show a statistically significant main effect proved unfounded. While the study has other problems, that is the way research is; even experts, using their best wisdom and experience as a base, can't always predict how it is going to turn out—and thank goodness! Granted, their record at prediction may be considerably better than chance, but still researchers must decide for themselves, occasionally disregarding expert opinion to make important new discoveries. We need risk takers, and we want to encourage those of you who are such. But we also want to warn the bulk of you that one can lose a great deal of time, effort, and resources going down this path as well. Think carefully when warned about your choice of problem.