2011 / 2012 Lesson Study Final Report:
Developing Students' Thought Processes For Choosing Appropriate Statistical Methods

Part I: Background

Authors: Elizabeth Knowles and James Murray, Department of Economics, University of Wisconsin - La Crosse

Disciplines: Research methods, statistics.

Submission date: August 23 2012.

Course Name: BUS 230: Business and Economics Research and Communication

Course Description: The course leads students through five main steps in the research process: the identification and definition of a research problem, research design, collection of the data used to study the research problem, statistical analysis of this data, and the interpretation and presentation of the results of the study. Students apply research methodology to a business problem in order to complete an original research project. The course is a core requirement in the College of Business Administration (CBA) undergraduate curriculum, and as such it must be taken by all students planning on majoring in any CBA major. The lesson focuses on the statistical analysis component of the course, specifically on helping students choose appropriate statistical procedures to answer their research question. Elementary statistics (MTH 145) is a prerequisite to the course, and the learning objective of this lesson builds upon and draws from the students' existing statistical knowledge.

Abstract: Introductory statistics classes typically emphasize computation and implementation procedures for a number of statistical tests. While it is essential to build these skills before achieving higher-order critical thinking skills, students often struggle in subsequent research methods courses when expected to select appropriate statistical tests to answer research questions. This requires an understanding of how statistical methods are related to one another; and to achieve this, students must develop a more advanced organization of knowledge. We designed a lesson to help students build a knowledge organization to achieve this outcome, and observed students to better understand their thought processes. We share our thought process map for selecting a statistical test, report on the impact it had for our students, and offer suggestions for improving the lesson. In addition, we describe the thought processes students used, both before and after being exposed to the thought process map, and identify sources of confusion revealed through the lesson study process. These include: when to apply an independent-samples test versus a paired-samples test, how the identification of scale of measurement led students to choose the wrong statistical method, the difficulty students had recognizing or defining what the variables in a problem were, and the lack of understanding of the difference between statistical language and colloquial language.

Part II: The Lesson

Learning Goal: Help students develop a way to organize knowledge of statistical tests that is conducive to applying this knowledge to answer research questions. To do so, we design a classroom lesson which presents a method for organizing knowledge of statistical tests, in addition to designing a set of activities which make learning evident, and an observation guide to record students' thought processes, before, during, and after this intervention.

Lesson Plan

Throughout the statistics portion of the course, we instructed our students on how to identify the appropriate statistical tests to answer research questions and used multiple examples to demonstrate the thought process. We both emphasized the consideration of the research question to be answered, the number of variables, the scale of measurement of the variables, whether observations were independent or paired, and whether the goal of the research question was to test for differences between groups or relationships/co-movement between variables. While students were exposed to a large number of statistical tests both in their introductory statistics classes and in the present business research methods class, the current lesson design focused on only the following four statistical tests: one-sample t-test, independent-samples difference in means test, paired-samples difference in means test, and a Chi-squared test of independence for two categorical variables. We conducted the lesson and made observations in both Professors Knowles' and Murray's classes. We did not make an explicit attempt to make the actual presentation of the material identical, so our students' performance may have been influenced by instructor-specific effects.

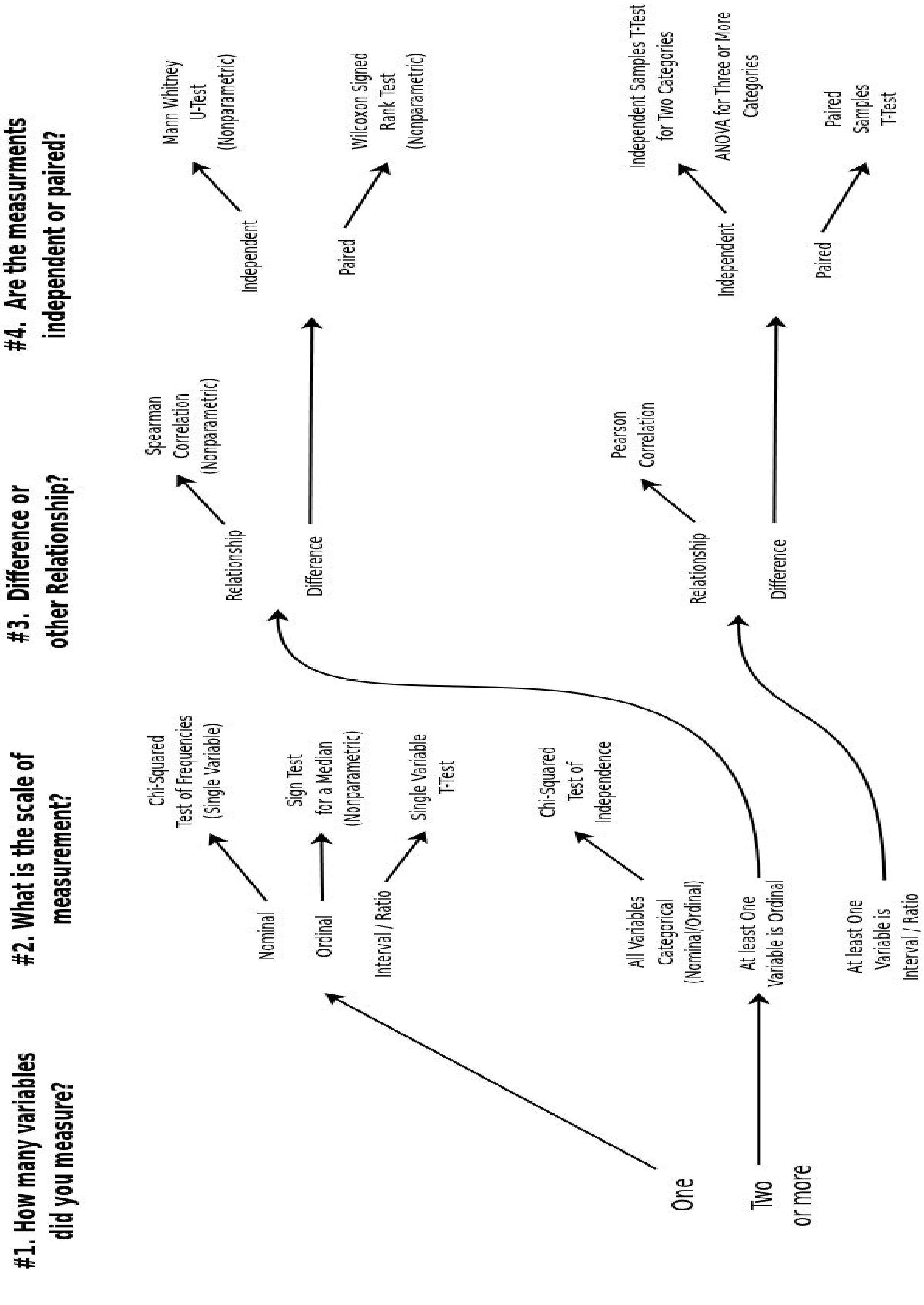
We jointly developed a single class period lesson on how to organize statistical knowledge, which included a brief lecture, the introduction of a statistical decision tree, and four in-class exercises (included in Appendix A). Each exercise presented a student with a description of a single research question and survey data which could be used to answer it. Students determined the appropriate statistical test, described aspects of the scenario that led them to this decision, and finally reported how confident they were in their answer. Students worked in groups on these exercises, but all students were expected to individually complete and turn in each exercise. We introduced the statistical decision tree as a way of organizing knowledge for students. It suggested the specific questions students should ask themselves when thinking about their decision and the order in which these questions are relevant. Although our classes differed from each other in the menu of statistical tests presented throughout the term (so each of our decision trees included a different set of branches and statistical tests), both included the same four statistical tests previously mentioned. Figure 1 shows a union of each of our decision trees, covering the statistical tests from both of our classes.

The lesson was delivered in two different ways in the Fall 2011 and Spring 2012 based upon when the statistical decision tree was introduced. In the Fall 2011 semester, students completed two of the in-class exercises at the end of the statistics unit. Then each instructor developed a statistical decision tree with the students to illustrate the thought process for the choice of statistical test. Following this, we instructed the students to put away their notes so that they could not reference the figure, and then complete the final two exercises. The intention of presenting the decision tree is to provide students a mental organization for their existing knowledge of statistical tests, not to provide a physical reference to depend on when tasked with choosing statistical tests. In the Spring 2012 semester, instead of presenting the decision tree at the end of the unit when the in-class exercises were being completed, we developed and used the decision tree throughout the entire unit on statistics. At the end of the statistics unit, students completed all four of the in-class exercises consecutively.

We conducted the lesson in multiple sections of the business research methods course, and each time we administered the four exercises in a different order in order to observe the effect presenting a knowledge organization had on choosing a statistical test, while accounting for a possible improvement in performance derived from simply practicing multiple exercises, as well as the possibility that students may simply be more familiar with some tests than others.

To measure how well students retained this knowledge organization, in the following week we gave our students an unannounced quiz and asked them to reconstruct the decision tree from memory.

Figure 1: Decision Tree



Part III: The Study

Approach

We observed our students' group discussions as they completed the four in-class exercises and noted whether the elements from the decision tree were considered in their discussions, and in what order they were considered. Our classroom observation guide is given in Appendix B.

We collected the students' written responses to the in-class exercises and noted whether or not the student correctly answered the question, how confident they were in their answer, and whether the students' logic included the following four considerations: 1) the number of variables involved in the research question, 2) the scale of measurement for the given variables, 3) the intent of the statistical test (i.e., whether the method tests for differences or a relationship between two variables), and 4) independent versus paired samples (relevant when considering either the independent samples t-test or paired difference t-test). If the students' responses did include one or more of these considerations, we noted whether their response to this consideration was correct for the given scenario.

Finally, we collected the decision trees that the students created from memory a week after the exercise took place. For each student we made note of whether they included each of the four considerations described in Figure 1, and whether or not they arrived to each of the four statistical tests without any incorrect considerations.

Findings / Discussion

IN-CLASS EXERCISE RESULTS:

Table 1 summarizes the percentage of students who correctly answered each in-class exercise for each presentation of the lesson. Each exercise is described in Appendix B, and are identified in Table 1 by the appropriate statistical test for each research question. The percentage of students correctly identifying each statistical test before and after introducing the decision tree I generally improves in Fall 2011. The improvement may be for two reasons: the students may simply improve after practice with each subsequent exercise, and/or the decision tree may have helped students better understand the statistical tests and given them a more effective mental organization from which to draw this knowledge. We examine this possibility more deeply below when we determine what considerations students used in their decision for each scenario.

The overall Fall 2011 percentage correct reveals the paired-samples t-test was the most difficult for students, as only 41% of students correctly identified this test. Interestingly, none of the 22 students who were given this question before the decision tree intervention answered this question correctly. Following the intervention, 67% of students answered it correctly.

The percentage correct for the Chi-squared test of independence shows that student performance actually decreased after the decision tree intervention. About 86% of students answered this question correctly before the intervention, and only 58% afterward. The most common incorrect answer for this question was a Pearson correlation coefficient, another test that examines a relationship between two variables, but used for interval/ratio data rather than categorical data. Classroom observations revealed that this decision was based on the word “relationship” appearing in the exercise, and even though many students correctly established that the scale of measurement was nominal, they still often made the wrong conclusion.

The results for Spring 2012 are somewhat disconcerting. The students had exposure to the decision tree for a few weeks in class while they were introduced to the statistical tests. Still, the results show that students performed as well as the the Fall 2011 students who had not yet been introduced to the decision tree, with the exception of the chi-square test in which students in Spring 2012 performed worse on average than in Fall 2011.

Table 1: Percent Correct for Each Statistical Test

	Fall 2011			**Spring 2012**
Statistical Test	**All Results**	**Before Decision Tree Introduced**	**After Decision Tree Introduced**	**All Results**
One Sample T –Test (A)	N = 57 79% Correct	N = 35 77 % Correct	N = 22 82% Correct	N = 84 92% Correct
Independent Samples T-Test (B)	N = 57 72% Correct	N = 35 54% Correct	N = 22 100% Correct	N = 84 56% Correct
Paired Samples T-Test (C)	N = 58 41% Correct	N = 22 0% Correct	N = 36 67% Correct	N = 83 47% Correct
Chi –Square (D)	N =58 69% Correct	N =22 86% Correct	N =36 58% Correct	N =84 29% Correct

Tables 2 through 5 describe the overall percentage of students that considered each of the following factors in their written responses: 1) the number of variables involved in the research question, 2) the scale of measurement for the given variables, 3) the intent of the statistical test (i.e., whether the method tests for differences or a relationship between two variables), and 4) independent versus paired samples. These are the considerations presented in the decision tree. In addition, the table also reports the percentage of students made the correct determination regarding the factor.

Table 2: One-Sample T-Test (A)

	Fall 2011			**Spring 2012**
	All Results (%)	**Before Decision Tree Introduced**	**After Decision Tree Introduced**	**All Results (%)**

	(N=57)	(N=35)	(N= 22)	(N=84)
Number of Variables was included/considered	90	83	100	96
Number of Variables was correct	75	71	82	90
Scale of Measurement was included/considered	68	57	86	71
Scale of Measurement was correct	63	57	73	69
Independent vs Paired samples was included/considered	14	6	27	4
Independent vs Paired samples was correct	86	94	73	96
Difference or relationship was included/considered	5	6	5	0
Difference or relationship was correct	95	94	95	100

Table 3: Independent-Samples T-test (B)

	Fall 2011			Spring 2012
	All Results (%) **(N=57)**	**Before Decision Tree Introduced (%)** **(N= 35)**	**After Decision Tree Introduced (%)** **(N=22)**	**All Results (%)** **(N=84)**
Number of Variables was included/considered	95	91	100	96
Number of Variables was correct	95	91	100	95
Scale of Measurement was included/considered	88	83	95	96
Scale of Measurement was correct	70	71	68	83
Independent vs Paired samples was included/considered	40	26	64	49
Independent vs Paired samples was correct	33	17	59	42
Difference or	61	49	82	60

relationship was included/considered				
Difference or relationship was correct	51	37	73	50

Table 4: Paired-Samples T-test ©

	Fall 2011			Spring 2012
	All Results (%) (N= 58)	Before Decision Tree (%) (N= 22)	After Decision Tree (%) (N= 36)	All Results (%) (N=83)
Number of Variables was included/considered	97	95	97	100
Number of Variables was correct	90	82	94	100
Scale of Measurement was included/considered	86	91	83	92
Scale of Measurement was correct	86	91	83	86
Independent vs Paired samples was included/considered	59	18	83	84
Independent vs Paired samples was correct	38	0	61	55
Difference or relationship was included/considered	33	50	22	80
Difference or relationship was correct	33	50	22	80

Table 5: Chi-Squared T-test (D)

	Fall 2011			Spring 2012
	All Results (%) (N=58)	Before Decision Tree Introduced (%) (N=22)	After Decision Tree Introduced (%) (N=36)	All Results (%) (N=84)
Number of Variables was included/considered	97	91	100	95
Number of Variables was correct	97	91	100	95
Scale of Measurement was included/considered	83	64	94	86
Scale of Measurement was correct	69	64	72	32

Independent vs Paired samples was included/considered	14	5	19	2
Independent vs Paired samples was correct	86	95	81	98
Difference or relationship was included/considered	62	82	50	77
Difference or relationship was correct	48	73	33	69

For the most part, students did consider the number of variables and the scale of measurement of the variables. Recognizing whether or not samples were independent or paired and whether the scenario suggested examining differences or a relationship occurred less frequently, and had mixed results by test. In Table 3 we see that the decision tree did lead to an improvement in this aspect of students' thought processes in Fall 2011. Before introducing the decision tree, less than half of the students (49%) considered whether a test for a difference or relationship was appropriate, and about one-forth (26%) considered whether the samples were independent or paired. Following the decision tree, the percentage of students considering difference versus relationship increased to 82% and the percentage of students considering independent versus paired samples increased to 64%. In Table 1, we see this is associated with an increase from 54% of students correctly answering this question before the intervention to 100% of students correctly answering this question after the intervention.

Table 4 reveals some consistent and inconsistent behavior regarding the paired samples t-test in Fall 2011. Following the intervention, students were more likely to consider independent versus paired samples, but were less likely to consider whether a test involved taking differences or looking for a relationship.

When comparing the before-decision-tree results in Fall 2011 to the results in Spring 2012 (when students had longer exposure to the decision tree) from Tables 2 through 5, in most cases students were more likely to consider each factor in their decision, and more often consider it correctly. Still, as is evident in Table 1, this failed to translate into a higher percentage of correct answers. We discuss these issues in the subsection below.

Tables 6 and 7 describe the students' presentation of the decision tree, drawn from memory one week following our lesson study classroom observation. Table 6 shows that in Fall 2011 almost all students included the number of variables and the scale of measurement in their decision trees. In this semester, only half of the students included the two, seemingly more difficult concepts of independent versus paired samples and whether a test is for a difference or relationship. We saw much better retention in Spring 2012. A very large majority of students remembered every important factor.

Table 6: Post Statistical Tree: Inclusion of elements

	Fall 2011		**Spring 2012**	
Trait Considered	Number of responses	Percent of Completed Trees	Number of responses	Percent of Completed Trees
Number of Variables	57	93%	82	99%
Scale of Measurement	57	89%	82	100%
Sample is Independent or Paired	57	54%	82	84%
Difference or Relationship	57	53%	82	89%

Table 7 describes the percentage of students that arrived to each statistical test without making any incorrect steps in the tree. to clarify, we noted that a branch was labeled as correct even if it was missing some considerations leading up to the statistical test, just so long as the considerations that were made were done correctly. On this basis, the students performed very well in Fall 2011, and extremely well in Spring 2012. In Fall 2011, it is evident that many students had not yet internalized the complete organizational structure of the decision tree, as many students' trees failed to include all four important factors, and many did not include all four statistical tests. There were some students, though, that were able to correctly reconstruct the entire decision tree from memory. We find that retention of the decision tree was far in Spring 2012. Most students included all four tests with no wrong information in the branch leading to the test. A large number were able to recreate without error the entire decision tree that was presented to them in class, which included many more tests than the four examined in this lesson.

Table 7: Post Statistical Tree: Test identified with no wrong information

	Fall 2011		**Spring 2012**	
Statistical Test	Number of responses which included test	Percent of Completed Trees that identified test with no wrong information	Number of responses which included test	Percent of Completed Trees that identified test with no wrong information
One Sample T –Test (A)	49	84%	80	94%
Independent Samples T-Test (B)	43	88%	69	93%
Paired Samples T-Test (C)	38	76%	66	92%
Chi –Square (D)	45	82%	75	93%

Considering our students' successes and failures in Spring 2012 relative to the students in Fall 2011 who had not yet been exposed to the decision tree, it is difficult to judge whether the statistical decision tree is an effective tool for helping students develop a thought process for choosing a statistical test. When measuring final performance in choosing the right statistical test to answer a question, the students did not do better, and often performed worse. However, in Spring 2012, students had high retention of the decision tree, the important factors behind choosing a statistical test, and the existing toolbox of statistical tests they had recently been exposed to. We expect students' failure in final performance using the decision tree derived from a lack of practice actually using it. In both Fall 2011 and Spring 2012, the exercises in this lesson were among students' first experiences actually using the thought process to make research design decisions, rather than it serving as a way to connect ideas in lecture. Given the successes we measured in Spring 2012, we expect we do have a valuable teaching tool, and that continued practice will not only make it more effective, but also closely integrate the classroom lecture on statistical tools with the thought process for choosing what tools to use in different situations.

CLASSROOM OBSERVATION:

In this subsection, we describe more about students' thought processes as evident from our experience sitting with the students and observing their conversations.

Students largely considered scale of measurement, both before and after the decision tree intervention. Often, scale of measurement was not stated so explicitly, but instead suggested in terms of what could be calculated with the data, such as a mean. Students often had difficulty with the concept of paired versus independent samples. It is unclear whether students did not consider this factor because they did not understand the difference, or because they did not understand the importance of this distinction in identifying the appropriate statistical test. One significant source of confusion concerned the word ``independent'' being used to describe samples with different observations versus having paired observations. Often they confused the word with another use of the word in the class for two variables being related to one another.

Following the decision tree intervention, many students used a process of elimination to come to a conclusion. Sometimes students eliminated a test because it had already been used to answer a previous question. Sometimes students used what appeared to be a quite random process of elimination, where they started with any test they could remember and tried to find reasons to eliminate it. More substantially though, some students used the process of elimination by considering a factor present in the decision tree and then eliminated tests on the basis of that factor. For example, sometimes students identified that a problem concerned a ratio variable and thereby eliminated the Chi-square test of independence. This is a reasonably good strategy, but the discussion indicated that there were factors in the decision tree that they were more comfortable with than others.

We also found that sometimes the decision tree caused more confusion, especially in Fall 2011 when students were initially introduced to the decision tree, then immediately asked to apply it. In some cases students asked the right set of questions, but many times were not able to answer these correctly, and were unable to reach the correct conclusion. When observing the order of the factors that students considered, in Fall 2011 it rarely followed the order that the tree presented, even after the decision tree was introduced. There was improvement in this regard in Spring 2012. In both semesters, we found that while students were trying to use the decision tree to choose a statistical test, they often did not pause to reflect on the purpose or intent of the research question first.

Both before and after the intervention, we identified other sources of confusion concerning statistical language. First is the use of the word, “independent” to describe samples with different observations in each sample. Secondly, students had trouble identifying what should be considered a variable. For the independent samples t-test, students were asked to determine whether the number of hours students spend studying (ratio variable) is different between students that are employed versus those who are not employed. Many times, students identified employment status as a nominal variable, which is appropriate. But when considering an independent samples t-test (the correct answer for this particular problem), some rejected the idea because that was a test for differences in means, which concerns two ratio variables.

Finally, students struggled with the colloquial use of the term, “relationship.” In the problem for the Chi-Squared tests of independence, students were asked to determine the relationship between employment status (given in three categories) and class standing (given in four categories). In both Fall 2011 and Spring 2012, many students peevishly hung on to that word and used it as a basis for deciding on a Pearson or Spearman correlation test.

CONCLUSION:

We found some limited evidence in Fall 2011 that teaching students the decision tree led to improved student performance and thought processes. Our classroom observations in both Fall 2011 and Spring 2012 revealed that students still did not completely understand the decision making process. The students’ decision trees drawn from memory in Fall 2011 revealed that they had not yet completely internalized the decision tree, even though it did lead to small improvements in student performance. The students’ decision trees drawn from memory in Spring 2012 revealed that there was strong retention of the decision tree, yet they had less success in implementing it.

We conclude that we do have a valuable lesson. Our lesson study revealed some unexpected sources of confusion, including understanding the difference between independent-samples versus paired samples, the implication of words such as “independence” and “relationship,” and what should be considered as a variable. As we proceed, we will strive to design class activities centered around the decision tree that build upon the successes revealed in Fall 2011 and Spring 2012.

Appendix: Lesson and Study Materials

<u>Appendix A</u>. In-class Exercise: Choosing Statistical Tools to Answer Research Questions

Directions: You will be given four research questions involving the survey questions below. Work in your groups to answer the questions when prompted. All papers will be collected.

Scenario: A researcher is interested in exploring the relationship between student employment and effort put forth towards academics. She administers a survey to full-time University of Wisconsin-La Crosse (UW-L) students which includes the following questions:

- What is your employment status? Full time / Part time /Not employed.
- On average, how many hours do you work per week? (Open ended)
- On average, how many hours do you study per week? (Open ended)
- What is your class standing?
  Freshman / Sophomore / Junior / Senior / Graduate Student /Other

1. Suppose the national average for the number of hours full-time college students work is 12 hours per week. The researcher is interested in determining if there is evidence that UW-L students study on average more hours than the national average.

2. The researcher is interested in determining whether there is a difference in the average number of hours students study per week between those who are employed (either full-time or part-time) and those who are not employed.

3. The researcher is interested in determining whether there is a relationship between class standing and employment status.

4. The researcher is interested in determining whether on average students spend more hours studying than the number of hours students spend working.

   a. What statistical method/test would you use to answer this question.

   b. Explain your reasoning for your answer in part (a). What characteristics of this research question and methodology make the test you chose appropriate.

   c. How confident are you that your answer to (a) and (b) is correct?

   Your level of confidence can be different than other members of your group.

   Very Confident **Somewhat Confident** **Not Confident** (Circle One)

Appendix B. BUS 230 Lesson Study Observation Guide

Observe the group discussion about the statistical test which is appropriate to answer the question. Record each observation with a number to indicate the order that the students consider each element. They may circle around to an element more than once, record this as it happens (i.e. any element may have more than one number beside it).

If the conclusion about the element is incorrect, record an X by the number.

Group number __________

Observation number (circle) 1 2 3 4

Element	Consideration
Reflect on the purpose or intent of statistical test (for example they talk about determining the difference or relationship).	
Discuss the number of variables considered	
Discuss the scale of measurement	
Discuss whether variables are independent or not	
Other	

Did students take into account any irrelevant considerations?

Did students reach the correct conclusion without well articulated reasons?

Did students reach the incorrect conclusion, but use mostly correct and well articulated reasons?

How many group members were actively engaged? (Include those who actively listening to understand the concepts, but not those just trying to write the correct answer)