A PRE-POST-RETENTION APPROACH TO ASSESS AND ASSURE LEARNING IN A BUSINESS STATISTICS COURSE AND BEYOND

Dr. David Roach
Professor of Management
Arkansas Tech University

Dr. Loretta Cochran
Asst. Prof. of Management
Arkansas Tech University

Introduction

Validation of learning has long been pursued by programs that desire to confirm the transfer of knowledge (Mirchandani, Lynch, and Hamilton, 2001). Moreover, pressure from accrediting agencies has led universities and business schools to increase their attention to assessment and assurance of learning. From the perspective of the university, the assessment continuum ranges from institutional measures to program measures to course measures. While each point on this continuum uses similar methodologies and are highly dependent, each level should have a different group of stakeholders (Huba and Freed, 2000). In this paper, we describe an approach to assess and assure learning as an extension of the assessment model - at the course level.

Assessment has traditionally been a top down process, with regional accrediting bodies directing institutions on self study requirements. From the university's perspective, accreditation – not assessment – holds the most value. A loss of accreditation can result in the inability for students to receive federal financial aid funds. On the other hand, if we are truly concerned about the assurance of learning, the place to start is course based assessment. Course level assessment linked to specific learning goals builds towards the more global assessment measures for a program or even an institution (Huba and Freed, 2000). In fact, the Education Commission of the States identifies "assessment and prompt feedback" as one of five components of quality instruction (1995, 1996). This is similar to the concept of building quality into a product rather than attempting to "inspect" your way to quality.

Though the assessments we describe in this paper are developed within the context of a specific course (Business Statistics), our approach also addresses program-related learning goals. For the statistics course itself, we consider the impact of the course on student learning as it relates to both (a) the technical knowledge (content) taught in the course and (b) the ability to think and reason statistically (del Mas, 2002). We then go on to consider the impact of the course on program-related goals, specifically the development of students' critical thinking and problem solving skills.

General Approach

To assess and assure course and program learning, we pose four questions:

1. "What do students know when they enter the course?"
2. "What do students know at various points in the course the course?"
3. "What do students know at the end of the course?"
4. "What do students retain from the course?"

The first three questions focus on learning as it develops in the semester the course is taught. Once answers to the three questions are obtained, assurance of learning can be addressed.

Comparing measures for the first two questions allows the instructor to monitor student knowledge/thinking and progress as the course unfolds. Midterm exams are one way to obtain information, but assessment can also occur daily. For example, Angelo and Cross (1993) present a variety of classroom assessment techniques (e.g., one-minute papers, muddiest point, empty checklists, etc.) to assess learning as the course progresses. These assessments permit the instructor to obtain ongoing feedback related to the presentation and student comprehension as the course unfolds. Then, if needed, adjustments (teaching methods, assignments, tutorials, etc.) can be made to assure student learning in the course.

Comparing measures for the first and third questions (pretest-posttest) allows the instructor to assess student knowledge/thinking at the beginning and end of the course. The difference between the pretest and posttest provides an indication of student learning in the course, but not without limitations. In particular, if questions are limited to course content (statistical literacy), then almost all variation in student responses at the beginning of the course is likely to be random. Moreover, students taking the pretest are likely to become frustrated and perhaps resentful. However, somewhat paradoxically, asking questions that address "higher-order" statistical reasoning and thinking may provide both a more meaningful and less frustrating approach. This deal will be developed further below.

Assuring learning in the course is necessary and vital, but, we contend, insufficient. For student learning to be assured, students must also retain and be able to apply that learning in the future, well after completing the course. The fourth question ("What do students retain from the course") directly assesses retention of course learning. Though there may be confounding variables (other courses, for example), comparing the third and fourth questions directly assesses the long term learning associated with the course. Table 1 summarizes the four questions, when they are asked, and what information they provide.

Table 1 – Research Questions and Measures Matrix

Question	Timing	Information
Q1 What do students know when they enter the course?	Pre-treatment measure	Baseline
Q2 What do students know at various points in the course?	Within Course Periodic Review	During treatment, indicates impact of presentation of material
Q3 What do students know at the end of the course?	Post -treatment measure	Knowledge/ability at end of course
Q4 What do students retain from the course?	Post Course Retention	Examination of knowledge at next application of course content

What Can and Should Be Assessed?

One dimension that can be assessed is *statistical literacy*, where statistical literacy involves students learning the basic technical knowledge in the class and is prompted by questions that ask students to identify, describe, rephrase, translate, interpret or read (del Mas, 2002). For example, a student might be (a) asked to compute the median from a set of numbers or (b) presented with a p-value of 0.001 and asked if the null hypothesis should be rejected (assuming a 0.01 level of significance). This would include the calculation of various descriptive statistics (measures of central tendency such as the arithmetic mean, measures of spread such as the standard deviation, and measures of association such as the correlation coefficient). In addition, this assessment would include topics related to various probability distributions (z, t, F, Chi-square, sampling distributions, etc.) and concepts and formulas related to inferential statistics (hypothesis testing and confidence intervals).

By contrast, many authors (Chance, 2002, del Mas, 2002, Melton, 2004, Rumsey, 2002) have raised questions about the focus on statistical literacy. These authors argue that most statistics courses have stated objectives that extend beyond statistical literacy. Moreover, despite these stated objectives, teaching methods and course examinations often fail to meet these objectives. As Chance (2002) puts it, this problematic because "you get what you measure." Expecting students learning to extend beyond what we teach and evaluate is wishful thinking.

Given the stated objectives for many statistics courses, a focus on statistical reasoning and statistical thinking is more appropriate. *Statistical reasoning* involves students learning to explain how or why results are produced and is prompted by questions that ask students to explain how or why a process works (del Mas, 2002). For example, a student might be asked to

(a) explain why the median is resistant to outliers or (b) explain why a low p-value should lead one to reject the null hypothesis.

Statistical thinking involves thinking about and identifying sources of variation (Melton, 2004), understanding how to obtain meaningful relevant data (Chance, 2002), developing a healthy skepticism about obtained data (Chance, 2002), and the application of the scientific method, i.e., identifying questions, formulating hypotheses, collecting and interpreting data, communicating results (Rumsey, 2002), involves students learning to "apply their basic literacy and reasoning in context" (del Mas, 2002) and is prompted by questions that ask students to apply, critique, evaluate or generalize (del Mas, 2002). For example, a student might be (a) asked to evaluate a short statement or passage such as: "According to a study commissioned by the governor's office, the average income in our state is higher than the average income in a neighboring state (p-value = 0.03)." A student who engages in "statistical thinking" would address issues such as the meaning of "average" and whether or not the study used the mean or median, how the data were collected, whether or not data were random and/or representative of the populations of interest, and the interpretation and meaning of the p-value.

Figure 1. Statistical Thinking (Chance, 2002; Melton, 2004; Rumsey, 2002)
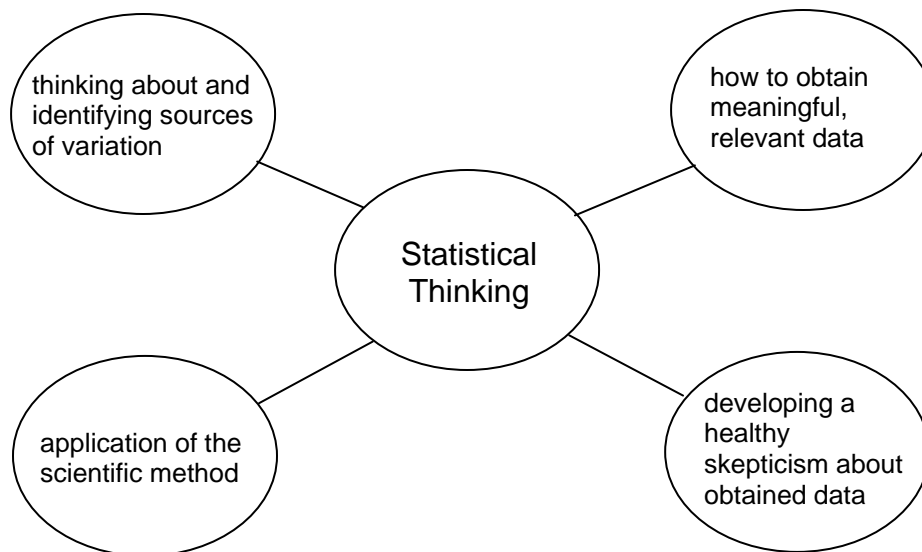
In addition to course-specific goals, statistical thinking can promote the development of students' *critical thinking* and *problem solving* skills. A student who thinks critically notices and asks questions, tries to use reason to answer those questions, and believes and internalizes the results of his or her reasoning (Nosich, 2005). For example, suppose a student reads a newspaper article claiming that taking vitamins is unhealthy. A student who thinks critically asks questions about the article, including questions concerning the evidence offered to support the claim asserted in the article and whether or not the evidence and argument provided in the article leads necessarily to the conclusion offered by the article's author. Thinking statistically, the student should

question the manner in which data were collected, sources of variation (including random variation and other plausible sources of systematic variation such as diet and dosage level) other than that presented by the author, and maintaining a skeptical attitude.

Problem solving involves recognizing and identifying a problem, gathering information, evaluating the information/evidence, consider alternatives and implications, and choose and implement the best alternative (Guffey, 1996). For example, the student might be presented with the following scenario:

> You manage the technical support group for a company that produces video games. This morning, your boss presented you with a customer letter complaining about the length of time spent waiting for "the next available customer representative" and the quality of technical advice provided. What should you do?

Thinking statistically, the student should question the nature of the problem (pervasive or isolated), how data/information might be gathered, and how the information/evidence should be evaluated.

Assessing "Introductory" Statistical Literacy

Broadly speaking, statistics can be divided into two broad categories: descriptive and inferential. The goal of descriptive statistics is to describe the key features of a data set using graphs and statistical summaries. Descriptive statistics include measures of central tendency (mean, median, and mode), proportions or percentages, linear relationships (slopes and y-intercepts) between variables, and various graphs such as histograms, box plots, bar charts, and scatter plots. Many students have at least some exposure to descriptive statistics. Topics such as central tendency and creating and interpreting graphs and tables are frequently covered in elementary school today. The notion of linear relationships should also be familiar, as almost all students taking a statistics class have completed at least one Algebra class. Some topics related to inferential statistics, such as margin of error, are at least somewhat familiar to most students because margin of error is almost always reported along with political and opinion polls pervasive in newscasts, especially in an election year.

For those statistics (or statistical concepts) students are likely to have encountered before the class, a pretest that assesses statistical literacy can be given early in the course, perhaps on the first day of class. In addition to providing a baseline for statistical literacy at the beginning of the course, the pretest can be used to ascertain the level of attention that should be directed toward particular topics. Questions addressing pre-course technical knowledge include:

1. In the following scatter plot, there is _____ relationship between house size and price.
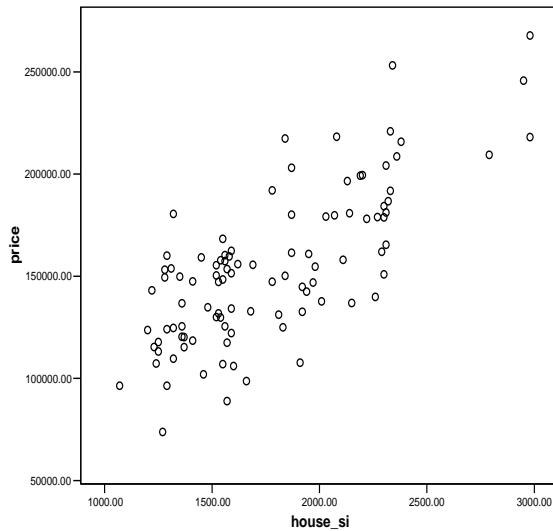
a. a positive
b. a negative
c. no apparent

2. Compute the grade point average for a student with the following course grades (A = 4 points, B = 3 points, C = 2 points, D = 1 point, F = 0 points).

Course	Credit Hours	Grade
Calculus	3	C
Biology	4	B
Economics I	3	C
Accounting I	3	B
Health and Wellness	2	A

a. 2.50
b. 2.77
c. 3.00
d. none of the above

Introductory statistical literacy can be assessed again at the end of the class, perhaps as part of a final exam. To the extent that there is a difference between the pretest and posttest, learning has occurred (assuming we have done no harm).

Introductory statistical literacy can then be assessed again some time after the course has been completed, perhaps as part of a pretest for a subsequent course. To the extent that there is a

difference between the statistics class posttest and the subsequent "retention" test, long-term learning has occurred. Currently, we are administering the retention test as a pretest in a Production Management course for which Business Statistics is a prerequisite. As students typically take the production class anywhere from one to three semesters after completing the statistics class, we will also, in time, be in a position to examine retention after various time spans.

Assessing "Advanced" Statistical Literacy and Statistical Reasoning

The goal of inferential statistics is to use statistics computed from sample data to estimate or test inferences concerning population parameters. Unlike descriptive statistics, most students entering the class have not been introduced to concepts of inferential statistics, with the possible exception of margin of error. As noted above, students are unlikely to have previously encountered this technical knowledge. As a result, variation in scores on a pretest is largely a function of random error. Moreover, if students are not knowledgeable about the statistical techniques, then they cannot be expected to explain why those techniques work; i.e., they cannot engage in statistical reasoning at this point. We recommend that advanced statistical literacy and statistical reasoning not be included on the course pretest. These components of learning should be examined during the course (exams, classroom assessment techniques) and as part of the final examination for the course (posttest). The posttest can also be administered as a pretest in the subsequent courses (Production Management in our case) to assess retention of material. Test items include:

1. A credit manager for a department store takes a random sample of 100 customer accounts and finds an average (mean) account balance of $75 with a standard deviation of 10. Assuming that the population distribution is a bell curve, the approximate limits for a 95% confidence interval for the population mean account balance are:
    a. 65 and 85
    b. 55 an 95
    c. 74 and 76
    d. 73 and 77

2. Following an advertising campaign for a new drug, the director of advertising for a pharmaceutical company wants to know if more than 50% of males between the ages of 50 and 70 are aware of the company's new drug. Based on a random sample of 500 males between the ages of 50 and 70, a p-value of 0.002 is obtained. With alpha set to 0.05, the director should conclude that:
    a. More than 50% of males between the ages of 50 and 70 are aware.
    b. Less than 50% of males between the ages of 50 and 70 are aware.
    c. Neither of the above is correct

3. In a random sample of 50 employees at *The Company*, the average number of absences over the past year is 6.4 with a standard deviation of 1.7. Using statistical software, the upper and lower limits for the 95% confidence interval for the population mean are 5.92 and 6.88, respectively. What does the 95% confidence interval tell you about the average high school percentile rank of freshmen at this university?
    a. Ninety-five percent of employees were absent between 5.92 and 6.88 days.
    b. There is a 95% chance that the true population mean is between 5.92 and 6.88.
    c. I am 95% sure that the true population mean is between 5.92 and 6.88.
    d. If we were to draw a random sample of size 50 over and over again, 95% of the confidence intervals would capture the true population mean.

4. As a check on service quality, management at the Better Computer Company measures the time that customers spend waiting for the "next available representative" when they call for technical assistance. In a random sample of 25 phone calls, the sample mean waiting time was 18 minutes with a sample standard deviation of 5 minutes. Testing the null hypothesis that the population mean is greater than or equal to 20 minutes, a p-value of 0.028 is obtained. What does this p-value tell you?
    a. There is a 2.8% chance that the population mean is 18.
    b. There is a 2.8% chance that the sample mean is 20.
    c. There is a 2.8% chance that the population mean is 20.
    d. If the population mean is 20, there is a 2.8% chance of obtaining a sample mean that is less than or equal to 18.
    e. If the sample mean is 18, there is a 2.8% chance that the population mean is 20.

Assessing Statistical Thinking Before After the Course

While students are unlikely to have been exposed to important technical statistical knowledge (particularly as it relates to probability distributions, sampling distributions, and inferential statistics) before the course, they are likely to have been exposed to the scientific method and other concepts related to statistical thinking. Paradoxically, the capacity to think statistically is not tied to specific technical knowledge and students are likely to arrive in the course with statistical thinking different abilities, and may employ scientific thinking (scientific method) or more general cognitive skills. Moreover, concepts such as the margin of error are frequently discussed on news programs describing political polls. Therefore, we propose testing including some items related to statistical thinking on the pretest, the posttest, and the retention assessment thinking. Items that examine statistical thinking does but do not *necessarily* require advanced statistical literacy include:

1. People who visit the website of a well-known polling organization are asked to vote on various issues. The question for last week was "Do you believe that there has been a general decline in corporate ethics in the past decade?" Sixty-four percent of the 12,000 respondents answered "yes." Based on these results, should you conclude that most (more than half) of U.S. citizens believe that there has been a general decline in corporate ethics in the past decade?
        a. yes   b. no

2. In 25 words or less, explain your answer to the preceding question.

3. In which of the data sets below would you be more confident that training method A leads to higher performance?
   a. Set A
   b. Set B
   c. neither (not enough information)

   Set A

Training Method A	2, 4, 7, 7, 10
Training Method B	2, 2, 3, 3, 10

   Set B

Training Method A	5, 5, 6, 7, 7
Training Method B	3,4, 4, 4, 5

4. In 25 words or less, explain your answer to the preceding question.

Assessing Critical Thinking and Problem Solving Skills

One overarching goal for a university education is to instill critical thinking and problem solving skills among students. Hence, in addition to examining course-related literacy, reasoning, and thinking we assess broader critical thinking and problem solving skills where statistical thinking can (and should) sharpen analysis. We ask questions in our pre, post, and retention tests. Examples include:

1. Discuss how well reasoned you find the argument presented below.

   Six months ago, *The Company* implemented a productivity improvement plan (PIP) at its Little Rock facility. Since that time, productivity has improved 10% at the facility. By contrast, in the Springfield facility where the PIP was not implemented, productivity has remained unchanged. Therefore, if *The Company* wants to raise productivity at all of its facilities, management should implement the PIP at all of its facilities.

2. Suppose you have to decide which of three training programs to implement. Describe the process you would use to make the decision.

Based on the critical thinking scale based on a scale developed by Faccione (1990), we develop a general scale to assess critical thinking skills. We also develop a question-specific scale for items that ask students to use statistical and critical thinking skills. We also developed a scale developed to assess problem solving. Appendix A includes all three scales.

Summary/Conclusion

In sum, we propose assessing various aspects of performance related to the statistics course at various points in time in order to arrive at a comprehensive assessment.   Though we focus on a specific course, we believe that a similar approach can be taken for other courses. We also contend that this approach can be used to focus teaching and assessment on the mission and objectives of higher education, namely a focus on reasoning and thinking that includes but goes beyond literacy within a specific course or courses. An overview of our approach and its application to the statistics course is provided in Table 2.

Table 2. Topic and Method of Assessment

What is assessed	When is it assessed	How is it assessed
Basic Statistical Literacy	Beginning of statistics course During statistics class End of statistics course Beginning of Operations course	Multiple choices, 10 item quiz – given in class.
Advanced Statistical Literacy	During statistics course End of statistics course Beginning of Operations course	Short multiple choice quiz given shortly after topic is covered in class and multiple choice quizzes given after course and at beginning of subsequent course
Statistical Reasoning	During statistics course End of statistics course Beginning of Operations course	Multiple choice and short answer questions that ask student to explain answers to multiple choice items
Statistical Thinking	Beginning of statistics course During statistics class End of statistics course Beginning of Operations course	Multiple choice and short answer questions that ask student to explain answers to multiple choice items
Critical Thinking, and Problem Solving	Beginning of statistics course During statistics class End of statistics course Beginning of Operations course	Evaluation of short written assignments

References

Chance, B. L. (2002), "Components of Statistical Thinking and Implications for Instruction and Assessment," *Journal of Statistics Education* [Online], 10(3). www.amstat.org/publications/jse/v10n3/chance.html

del Mas, R. C. "Statistical Literacy, Reasoning, and Thinking: A Commentary" *Journal of Statistics Education* [Online], 10(2). http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html

Education Commission of the States, (1995). *Making quality count in undergraduate education.* Denver, CO: Education Commission of the States.

Education Commission of the States, (1996, April). "What research says about improving undergraduate education." *AAHE Bulletin*, 5-8.

Huba, M E. and Freed, J. E. (2000). *Learner-Centered Assessment on College Campuses*. Boston: Allyn and Bacon.

Melton, K.I. (2004). "Statistical Thinking Activities: Some Simple Exercises With Powerful Lesson," *Journal of Statistics Education* [Online], 12(2). www.amstat.org/publications/jse/v12n2/melton.html

Nosich, G.M.,(2005). *Learning to think things through: A guide to critical thinking across the curriculum*, Upper Saddle River, NJ: Pearson Education.

Paul, R. & Elder, L. (2002). *Critical Thinking: Tools for taking charge of your learning and life*, Upper Saddle River, NJ: Pearson Education.

Rumsey, D.J. (2002). "Statistical Literacy as a Goal for Introductory Statistics Courses," *Journal of Statistics Education* [Online], 10(3). www.amstat.org/publications/jse/v10n3/rumsey2.html

Wild, C. J. (1994), "Embracing the 'Wider View' of Statistics," *The American Statistician*, 48, 163-171.

APPENDIX A

Statistical Thinking for the PIP passage
1. Thinking about and identifying sources of variation
    [0-Not present, 1-present but weak, 2-present and strong]
    -Recognized other possible internal causes of performance differences
    -Recognized other possible external causes of performance differences
2. Understanding how to obtain meaningful relevant data
    [0-Not present, 1-present but weak, 2-present and strong]
    -Distinction between productivity and quality
    -Measurement of productivity
3. Developing a healthy skepticism about obtained data
    [0-Not present, 1-present but weak, 2-present and strong]
    -Question credibility of the author
    -Question whether or not people at plant were influenced by just being observed
4. Application of the scientific method
    0-Not present, 1-present but weak, 2-present and strong]
    -Not applicable in this example

Critical Thinking
    1. Accurately interprets evidence, statements, graphics, questions, etc.
       [0-Not present, 1-present but weak, 2-present and strong]
    2. Identifies the salient arguments (reasons and claims) pro and con.
       [0-Not present, 1-present but weak, 2-present and strong]
    3. Thoughtfully analyzes and evaluates major alternative points of view.
       [0-Not present, 1-present but weak, 2-present and strong]
    4. Draws warranted, judicious, non-fallacious conclusions.
       [0-Not present, 1-present but weak, 2-present and strong]
    5. Justifies key results and procedures, explains assumptions and reasons.
       [0-Not present, 1-present but weak, 2-present and strong]
    6. Fair-mindedly follows where evidence and reasons lead.
       [0-Not present, 1-present but weak, 2-present and strong]

Problem Solving
    1. identifying a problem
       [0-Not present, 1-present but weak, 2-present and strong]
    2. gathering information
       [0-Not present, 1-present but weak, 2-present and strong]
    3. evaluating the information/evidence
       [0-Not present, 1-present but weak, 2-present and strong]
    4. consider alternatives and implications
       [0-Not present, 1-present but weak, 2-present and strong]
    5. choose and implement the best alternative
       [0-Not present, 1-present but weak, 2-present and strong]