Paper presented at the ARTIST Roundtable Conference on Assessment in Statistics
held at Lawrence University, August 1-4, 2004

Anticipating and Addressing Student Misconceptions

Allan J. Rossman and Beth L. Chance
Department of Statistics
Cal Poly – San Luis Obispo
arossman@calpoly.edu, bchance@calpoly.edu

Introduction

Much of statistics is based on the fact that while individual values vary, their pattern of variation in the aggregate is often quite predictable. Similarly, individual students of statistics vary greatly not only in their developing levels of understanding but also in the gaps in their understanding and the errors that they make. Nevertheless, considering students in the aggregate, we can make some reasonable predictions about common misunderstandings and errors that students will make as they learn our subject. These problems occur in areas of performing calculations, understanding concepts, and presenting results. For example, we know that some students will fail to arrange the data values in order before calculating a median and that some will ignore the role of sample size when calculating a standard error. We know that some students will confuse "bumpiness" with variability and that some will misinterpret a confidence interval as a prediction interval. We know that some students will fail to label axes in a scatterplot and that some will leave a conclusion at "reject the null hypothesis" without relating the conclusion back to the context. The sooner we can recognize, diagnose, and address such misconceptions, the more effectively we will be able to help students learn statistics. We must also consider that many students enter introductory statistics courses with low confidence and motivation, which also impacts the type of assessments that will be effective and informative to both the students and the instructors.

In this article we describe two types of informal, formative assessment items that we are developing: "What Went Wrong?" exercises and "Practice Problems." The "What Went Wrong?" exercises were developed as part of a student manual to accompany *Workshop Statistics*. The primary goal is to maintain the constructivist approach of the main text by presenting students with sample responses that contain some kind of error and asking students to identify and correct the errors. We hope these exercises will improve students' ability to evaluate statistical arguments and to anticipate common pitfalls to be aware of. This should enhance their own problem solving skills and improve their ability to assess the reasonableness of their own answers. The "Practice Problems" have been embedded in a post-calculus introductory statistics course to supplement more standard homework problems. We envision these exercises as short initial reviews, applications, and extensions of the terminology and concepts presented in the preceding class session. Students complete and submit their solutions between classes, allowing the instructor to quickly get a sense for where student (both individually and collectively) understanding is weak and to tailor subsequent class discussion based on these responses.

In this paper, we further describe our goals for these assessments and briefly suggest the principles on which they are based. Then we present several illustrative examples of the assessment items. We conclude with some preliminary evaluation results, although this is very much work in progress and we look forward to feedback from the participants in this conference.

Goals and Principles

While these assessment approaches were initially developed for two different student audiences, they share some common goals:

- To provide students with timely feedback on their understanding of important ideas and skills in a non-threatening, supportive environment;
- To develop a feedback loop between the instructor and students that allows the instructor to determine the needs of students and tailor instruction to meet those needs.

Some principles on which this assessment strategy is based include:

- *Learning occurs through struggling and wrestling with ideas* (e.g., Resnick, 1987). With these assessments, we are trying to provide a low-pressure, engaging environment in which students test their understanding but with sufficient structure to provide meaningful feedback to the students. In this way, the assessments become embedded in the instructional process.
- *Addressing misconceptions head-on is a crucial step in the learning process* (e.g., Garfield, 1995). Student learning is enhanced when they are forced to confront their own misconceptions. We have developed these assessment items based on what we have seen as the most common errors made by our introductory statistics students over a combined 25 years of teaching experience and also from classroom-based research. These assessments are designed to provide students with feedback much more quickly than waiting for a midterm exam or even for a homework assignment.
- *Developing students' abilities to critique others' arguments is worthwhile for its own sake, and the practice of explaining errors in others' reasoning can be a powerful way to deepen students' understanding* (e.g., Pressley and McCormick, 1995). These exercises have the advantage of not requiring the student to have the misconception or make the error in the first place. This expands the number of misconceptions that students will encounter, and we believe it is less intimidating for them to correct other student work, encouraging them to learn from their own and others' errors.
- *Provide immediate and constructive feedback.* The strategy of "just-in-time teaching" (Novak et. al., 1999) is to provide a feedback loop formed by students' out-of-class preparation that impacts what the instructor presents in class. This strategy aims to maximize the effectiveness of in-class sessions based on students' understandings and needs, and also to structure students' out-of-class preparation time for maximum learning benefit. We have designed these assessments to better tailor how students spend their time outside of class, to help highlight to them the most important concepts to be remembering from each lecture period, and to provide suggestions for how students can improve their understanding.

In the following we provide examples of the assessment items we have been developing, highlighting their intending impact. While we have not yet used the student guide, we have been utilizing some of these questions for several years in our algebra-based statistics courses. The "Practice Problems" were used last spring with more mathematically inclined students. These are intended as "out of class" assessments but which impact classroom discussions.

"What Went Wrong?" Exercises

The following assessment items are taken from Chance and Rossman (2005a).

1) The following table presents data on the "number of inversions" made by 144 roller coasters during one run, as obtained from the Roller Coaster DataBase (www.rcdb.com).

Number of inversions	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Tally (count)	*84*	*7*	*12*	*11*	*9*	*9*	*3*	*9*

Consider the following possible solutions for calculating the mean and median. Identify what the student did incorrectly in each case.
(a) A fellow student reports the median to be 3.5 inversions.
(b) A fellow student reports the mean to be 18 inversions.
(c) A fellow student reports the median to be 9 inversions.
(d) A fellow student reports the median to be 0. (The problem here is an error of presentation, not calculation.)

The misunderstanding in (a) is a common one: it arises from treating the 8 categories (the integers 0-7) as 8 observations and applying the $(n+1)/2$ rule to find that the location of the median is between the 4th and 5th observations. Students need to recognize that the observational units here are the roller coasters, so n=144. The median therefore occurs between the 72nd and 73rd observations, which the tallies reveal to both equal zero, so the median is zero inversions. The errors in (b) and (c) both result from a different common misunderstanding: treating the tallies (counts) as the data. The mean of the eight tally values reported in the table (84, 7, 12, 11, 9, 9, 3, 9) equals 18, and the median of them equals 9. Students should notice that both of these values are clearly impossible, because the table reveals that the largest number of inversions in any coaster is 7. The error in (d) is quite minor but worth noting: in presenting the median value the student should report the units (the median is 0 inversions) rather than the naked number.

2) Suppose that a student Carlos is asked to investigate whether there is a relationship between whether he has seen a movie on the American Film Institute's "top 100" list and whether his lab partner Hannah has also seen it. Suppose that Carlos reports back that he has seen 36 of the movies and Hannah has seen 27. What's wrong with his presenting the table of data as:

	Carlos	*Hannah*
Yes	*37*	*27*
No	*63*	*73*
Total	*100*	*100*

This is a common error that students make when creating two-way tables from raw categorical data. This table only gives us the marginal information (Hannah's outcomes and Carlos' outcomes) and does not allow us to calculate conditional proportions to see if the two variables ({has Hannah seen the movie?} and {has Carlos seen the movie?}) are related. Students should note immediately that the table has a grand total of 200 movies, which indicates that something is amiss. To properly construct the two-way table, we need to put one variable as the column variable and one as the row variable. All we can fill in based on the given information is the margins of the table:

	Carlos: yes	Carlos: no	Total
Hannah: yes			37
Hannah: no			63
Total	27	73	100

3) Suppose that a student is given the following problem:

> *In the past, about 10% of people were left-handed. You suspect the percentage is increasing as people are less likely to dissuade a child from being left-handed these days. You observe a random sample of 100 first-graders a recess to see if they throw with their left or their right hand and find 15 use their left hand. What is the probability that you would observe 15 or more first-graders using their left hand if the population proportion was still 10%?*

Identify the errors in the following calculations:

(a) $z = (15 - .10)/(.10/\sqrt{100}\,) = 1490$; *probability above* ≈ 0

(b) $SD = \sqrt{.1(.9)/100} = .03$; $z = (.15 - .10)/(.03/\sqrt{100}\,) = 16.67$; *probability above* ≈ 0

The response in (a) contains at least two errors: the student has not converted the count of 15 into a proportion, and the standard deviation calculation in the denominator is also incorrect. We have seen students make this mistake because they do not recognize that this question is about a categorical variable and therefore a proportion, so they mistakenly try to calculate a *z*-score for a sample mean. On the other hand, the response in (b) has correctly calculated the standard deviation of the sample proportion, but then it mistakenly divides that standard deviation by the sample size 100 in the denominator of the *z*-score calculation. Perhaps students make this error because they again confuse the formulas for the standard deviation of a sample proportion and of a sample mean.

4) Suppose that 36 students report how much sleep they got on the previous night, with a resulting sample mean of 6.24 hours and a standard deviation of 1.34 hours. Identify the errors in the following calculations and interpretations of a 90% confidence interval for the population mean:

(a) The confidence interval for μ *is* $6.24 \pm 1.645(1.34/\sqrt{36}\,)$

(b) The confidence interval for μ *is* $6.24 \pm 1.69\ (1.34)$

(c) We are 90% confident that the sample mean falls between 5.86 and 6.62 hours.

(d) We are 90% confident that if we take a new random sample, its sample mean sleeping time will fall between 5.86 and 6.62 hours.

(e) 90% of the students in the sample slept between 5.86 and 6.62 hours

(f) 90% of the students in the population slept between 5.86 and 6.62 hours.

(g) Because $n \geq 30$*, we can assume that the population follows a normal distribution*

The response in (a) contains the minor error of using a *z*-critical value rather than a *t*-critical value even though the interval is based on the sample and not the population standard deviation. Response (b) has the more important error of forgetting to divide the sample standard deviation by the sample size in order to determine the standard error of the sample mean. The interpretations in both (c) and (d) mistakenly state that the interval estimates a sample mean rather than a population mean. The statements in (e) and (f) confuse a confidence interval with a prediction interval, as they claim that most students in the sample, or in the population, will have a sleep time in the interval. Finally, response (g) makes the common error of asserting that the large sample size makes the population distribution normal, rather than recognizing that the large sample size will make the sampling distribution of the sample mean approximately normal, which justifies the use of the *t*-interval.

Discussion
Preliminary feedback from reviewers of the student guide indicates that these exercises are addressing common misconceptions and that they maintain the constructivist nature of the main text. However, there is some concern that students may not always see the error. In fact, we do not expect students (or even experienced instructors) to identify them all correctly, and their time would not be well spent in trying too hard to decipher the error. What we do consider important and valuable is the student learning that can result from the discussion that follows these exercises. This discussion, which can be either an in-class discussion or simply a written discussion in the book, can help students to consider the concept in a different way and thus deepen their understanding. By attempting to find these errors, students will increasingly develop the important habit and skill of checking the reasonableness of an answer before proceeding. The experience also helps to put students in the mode of "teaching to others" which should also strengthen their understanding. Both the instructor and the student can learn much about the student's level of understanding from his/her explanations of other students' wrong answers.

From these examples, we have tried to illustrate that the types of problems we ask students to consider include calculation details, interpretations and conceptual understanding, and even presentation of results. This illustrates to students that all of these "skills" are important to master and also provides insight to the students as to how the instructor will be evaluating their responses (e.g., missing context). This can be especially powerful when the student either makes the same error or does not see a way to improve the response initially. By providing the errors and asking the students to critique someone else's work, when they correct an error it serves as positive reinforcement instead of students feeling that the instructor is always being needlessly picky or only providing negative feedback.

"Practice Problems"
The following assessment items are taken from Chance and Rossman (2005b).

Practice 1-1) Go take a hike!
The book Day Hikes in San Luis Obispo County *by Robert Stone gives information on 72 different hikes that one can take in the county. For each of the 72 hikes, Stone reports the distance of the hike (in miles), the anticipated hiking time (in minutes), the elevation gain (in feet), and the region of the county in which the hike can be found (North County, South County, Morro Bay, and so on for a total of eight regions).*
(a) What are the observational units here?
(b) How many variables are mentioned above?
(c) Classify each variable as quantitative or categorical.

(d) If you create a new variable that is the ratio of hiking time to hiking distance, would that be a quantitative or categorical variable?
(e) If you create a new variable that is coded as "short" for a hike whose distance is less than 2 miles, "medium" for a hike whose distance is at least two but not more than four miles, and "long" for a hike whose distance is at least four miles, would this be a quantitative or categorical variable?
(f) For each of the following, specify whether or not it is a legitimate variable for the observational units you specified in (a).

- *Longest hike in the book*
- *Average elevation gain of a hike in the book*
- *Whether or not the hike is in the North County region*
- *Proportion of hikes with an elevation gain of more than 500 feet*
- *Anticipated hiking time, reported in hours*

(g) Consider the 50 states as the observational units for a study. Suggest one quantitative and one categorical variable that you could measure about the states. (Be careful to express your answers clearly as variables.)
(h) Consider the 9 members of the Supreme Court as the observational units for a study. Suggest one quantitative and one categorical variable that you could measure about the justices. (Be careful to express your answers clearly as variables.)

This problem concerns the basic ideas of observation/experimental unit and variable that permeate the entire course. We have come to believe that students' lack of understanding these fundamental ideas is a large obstacle to their understanding more complicated ideas later in the course. For example, they can not understand a sampling distribution, where the observational unit becomes the sample and the variable is the value of the sample statistic, if they do not have a firm understanding of the basic terms. Therefore, after defining these terms and describing types of variables (categorical, quantitative) very early in the course, we ask students to apply those definitions to make sure that they understand them well.

Practice 2-13) More on sleep deprivation
(a) Recall the difference in group medians between these two groups.
(b) Change one line of your Minitab macro from Investigation 2-8 so that it analyzes difference in group medians rather than differences in group means. Apply your macro to the original data from the sleep deprivation study, using at least 1000 repetitions. Describe the distribution of differences in group medians, and report the empirical p-value.
(b) Comment on how the randomization distribution of the differences in medians differs from your earlier analysis of means.
(c) What conclusion would you draw about the statistical significance of the observed difference in group medians from this analysis? Explain.
(d) Now apply the macro for analyzing differences in medians to the hypothetical data from Investigation 2-9. Again report the approximate p-value, and compare this analysis to your earlier analysis of means.
(e) Does your earlier conclusion that the difference is much less significant for the hypothetical data than for the real data hold for analyzing medians as well? Explain.

This problem asks students to conduct a straight-forward extension of an analysis that they just completed in class. They return to a dataset for which they wrote a Minitab macro to simulate a randomization test for comparing the mean responses between two groups. In the practice problem students adjust the macro to simulate the randomization distribution for the difference in medians between the groups, and they comment on how their conclusion differs depending on whether they analyzed means or medians. Not only does this practice problem gauge students' ability to conduct and interpret the simulation, it also generates discussion about the trade-offs between analyzing the mean or median as a measure of center.

Practice 4-22) More on mothers' ages
Suppose you obtain a sample mean of 22.52 years and sample standard deviation 4.88 years and plan to construct a 95% confidence interval for μ, the population mean age of mothers at first birth.
(a) What additional information do you need to construct the confidence interval?
(b) Calculate a 95% confidence interval for n=5, n=25, and n=1199.
(c) How do the intervals in (b) compare? Explain why this makes sense.
(d) Suppose the sample standard deviation had been 9.75 years instead. Recalculate the 95% confidence interval for n=25. How does it compare to the interval with s=4.88? Explain why this makes sense.
(e) Calculate a 99% confidence interval for n=25. How does it compare to the interval with C=95. Explain why this makes sense.
(f) Suppose the sample mean had been 32.52, n=25, C=95%. How does the interval change? Explain why this makes sense.

This practice problem immediately follows students' first experiences with *t*-intervals for estimating a population mean. It aims to test students' understanding of the effects of various elements (sample size, sample mean, sample standard deviation, confidence level) on the *t*-interval. Students would have already encountered the effects of sample size and confidence level when they earlier studied confidence intervals for a population proportion. They also get some experience with calculating *t*-intervals, either by hand or as we would suggest with technology. Asking students to complete this outside of class frees up class time for the instructor to focus on areas where students struggle, rather than devote considerable class time to topics that students are able to understand on their own.

Discussion
Even in a very interactive classroom environment, both the instructor and the students can be fooled into believing that students' levels of understanding are higher than they really are. The true test of their understanding comes when students leave class and try on their own to apply that they have learned. These practice problems force students to test that understanding immediately, before the next class period, but again in a non-threatening environment. We grade these problems more on participation than exactness, so students are encouraged to attempt them but with the focus on learning instead of an assigned grade. We provide feedback immediately through written comments on individual student responses, through discussion at the beginning of the next class session, and/or through the course delivery software (the instructor entered both the questions and the "answers" which were delivered to students after they submitted their responses; further feedback can be tailored to the individual student response given). In many

cases there is not a definitive answer, but the problem can serve to foster classroom discussion and debate. This also reinforces to students that in the practice of statistics, it is important to continue to learn and to continue to evaluate alternative approaches.

In designing the "Practice Problems," we sought to focus on the most important and most problematic areas for students. This is especially helpful for allowing sufficient time to ensure student understanding of fundamental ideas like "variable" and effect of sample size, which recur often and underlie other important ideas and techniques. Thus, the problems help students prioritize the material that they studied in that previous class period and to identify the most important "take-home" knowledge. Sometimes the practice problem is a reapplication of a previous analysis and sometimes it is the "next step" we want students to take in their understanding or further exploration of a key concept. Either way, the problems are designed to be "small steps" to increase the likelihood of students being able to apply their new knowledge and to reinforce that knowledge, and also to build their confidence. Accordingly, the instructor will probably want to supplement these Practice Problems with more open-ended homework problems and other more authentic assessment tasks where the solution approach is less well-defined by the instructor. By structuring this continual review for the students, it requires them to consider the material several times during the week instead of only prior to a homework assignment due date. (We have found this approach to studying statistics, akin to how one might approach learning a foreign language, to be very effective and not one students might naturally consider. See also Garfield, 1995.)

It is also still important for instructors to review and respond to the student responses. The practice problems allow this to be done more informally, enabling the instructor to gain a quick snapshot of where the students are in their understanding before proceeding. Converting the problems to multiple choice items and using a course management system to record the frequency of responses can further expedite this process. Still, we also encourage the additional insight that can be gained from asking the students to explain their understanding in their own words.

Preliminary Evaluation

The Practice Problems are an important component of Chance and Rossman (2005b), the preliminary version of a new text written to support an active learning approach to studying introductory statistics at the post-calculus level. We class-tested these materials in the spring of 2004 at St. Olaf College and at Cal Poly. When students were asked whether the Practice Problems helped them to learn the course material, 23 of 29 students (79.3%) either agreed or strongly agreed. On the issue of whether these Practice Problems should be required, students were less in agreement: 17 of 29 (58.6%) agreed or strongly agreed with that assertion. Individual comments on open-ended questions of the course evaluation revealed that some students appreciated the role of the Practice Problems in developing their understanding:

- "The practice problems, I felt were a necessity because otherwise I think I would have not taken the time to actually do the problems in the book and been completely lost on what was going on."
- "I liked the format of the practice problems that we had this semester…. It was helpful that they were not graded because there wasn't as much pressure. But since they were

part of the participation grade, we still had to spend time on them and spending time on the practice problems made the discussions much more helpful."

- "I like the fact that practice problems are required but not graded, it takes the pressure off of something if you don't understand it at first."

Conclusions

We propose that these two types of assessments can help students' learning. The "What Went Wrong?" exercises force students to confront and explain common misconceptions, whether or not they suffer from that particular misconception themselves. The "Practice Problems" can help the instructor to diagnose areas in which students' understandings are weak, in order to devote more class time to addressing those misunderstandings. Both types serve as informal, formative assessments that engage students, are designed to reveal and confront common misconceptions, and aim to deepen students' understanding through positive reinforcement in a comfortable environment. We believe such assessment tools are very valuable in helping both the student and the instructor to much more quickly pinpoint student errors, correct misunderstandings, and reinforce important ideas.

References

Chance, B. and Rossman, A. (2005a), *Workshop Statistics Student Toolkit*, Emeryville, CA: Key College Publishing

Chance, B. and Rossman, A. (2005b), *Investigating Statistical Concepts, Methods, and Applications* (preliminary edition), Pacific Grove, CA: Duxbury Press.

Garfield, J. (1995), "How Students Learn Statistics," *International Statistical Review,* 63, 25-34.

Novak, G., Patterson, E., Gavrin, A., and Christian, W. (1999), *Just-in-Time Teaching: Blending Active Learning with Web Technology*, Upper Sadlle River, NJ: Prentice Hall. See also http://webphysics.iupui.edu/jitt/jitt.html.

Pressley, M. and McCormick, C. (1995), *Advanced Educational Psychology for Educators, Researchers, and Policymakers*, New York: Harper-Collins.

Resnick, L. (1987), *Education and Learning to Think,* Washington, D.C.: National Research Council.