Reasoning and Communicating in the Language of Statistics

Carol S. Parke
Duquesne University

Key Words: conceptual understanding, confidence, interpretation of results, verbal communication, written communication

Abstract

Although graduate students in education are frequently required to write papers throughout their coursework, they typically have limited experience in communicating in the language of statistics, both verbally and in written form. To succeed in their future careers, students must be provided with opportunities to develop deep understandings of concepts, develop reasoning skills, and become familiar with verbalizing and writing about statistics. The instructional approach described here spans the entire semester of a statistics course and consists of several aspects including cognitively rich individual assignments, small group activities, and a student-led scoring activity. To demonstrate the impact of this approach on student learning, qualitative and quantitative data were collected from students in two statistics courses. Several assessments indicate improvement in students' reasoning and understanding, written and verbal communication, and confidence.

1. Introduction

A reform movement has emerged over the past two decades within the teaching and learning of statistics. Instructional emphasis is shifting from rote memorization of formulae, computational skills, and procedural rules to conceptual understanding, making connections among statistical concepts, use of real-world data, interpreting results, and making appropriate conclusions. Educators and researchers use terminology such as statistical reasoning, thinking, comprehension, citizenship, and literacy to describe desirable goals of instruction and assessment. As early as the 1970s, researchers wrote about statistical reasoning. Chervaney, Collier, Fienberg, Johnson, & Neter (1977) presented a three-step process of reasoning that focused on comprehending the problem, planning and executing appropriate methods to solve the problem, and evaluating and interpreting the outcomes. However, the early research did not describe how to operationalize these ideas in the statistics classroom.

Recent literature defines student learning outcomes and describes instructional techniques that promote a

solid understanding of statistics. A series of articles published in 2002 discussed the varied definitions of three domains of statistical development (literacy, reasoning, and thinking) and provided illustrations of ways to incorporate the domains into the teaching, learning, and assessment of statistics. One perspective considers the three domains to be mainly independent of each other while recognizing some degree of overlap. Another perspective views statistical literacy as the overarching goal serving as the umbrella under which reasoning and thinking are subsumed (delMas, 2002).

Garfield (2002) describes statistical reasoning as making sense of statistical information, making inferences, interpreting results, and having a conceptual understanding of important ideas. She presents a five-stage model of statistical reasoning for sampling distributions that progresses from faulty reasoning and misconceptions to integrated process reasoning in which "the student has a complete understanding of the process of sampling and sampling distributions, …can explain the process in his or her own words, and makes correct predictions with confidence" (p. 7).

Statistical thinking is described by Chance (2002) as the ability to view the statistical process as a whole, to explore data in ways beyond those given in textbooks, to pose new questions, to understand the meaning of variation, and to understand "why". She provides illustrations of how to instill these mental "habits" in students during instruction. They include examining data collection issues to ensure that the right data is collected to answer the right questions, providing opportunities to be skeptical of statistical results in the media, examining limitations of inferential techniques, relating the data to the context, and understanding the relevance of statistics.

Finally, statistical literacy is defined by Rumsey (2002) as encompassing two distinct learning outcomes: competence and citizenship. Statistical competence is the basic knowledge underlying statistical reasoning and thinking. Statistical citizenship is the goal of becoming educated consumers of quantitative data who are able to think critically and make good decisions from the information. Rumsey also states that correctly calculating numerical results does not necessarily demonstrate a good understanding of concepts. Students should be given opportunities to collaborate with each other and to explain ideas and interpret results in their own language.

1.1 Written and Verbal Communication in Statistics

An essential aspect of statistical literacy is the ability to communicate concepts and results in written and verbal form. In her article on literacy, Rumsey (2002) distinguishes between interpretation skills that demonstrate whether a student understands a concept and communication skills that involve sharing the information clearly with others. In the late 1980s and early 1990s, much of the research on writing in statistics reflected the "writing across the curriculum" movement in colleges and universities. Instructors in all programs were required to incorporate writing into their courses. In statistics classes, the requirement was often met by adding a writing component to the course. The component typically asked students to write a paper about a topic learned in class, complete a data analysis project, summarize a method, or respond to essay items on an exam (e.g., Iversen, 1991; Peterson, 1988). Writing was not integrated throughout the course, but rather it was an "add-on". Asking students to write about statistics in a paper, on an exam, or in an assignment, does not automatically increase their understanding or improve their communication skills.

Altering the structure of a statistics course is one way that some instructors have successfully incorporated writing. Spurrier (1997, 2000) describes a one-credit capstone course for statistics majors. Students are placed in the role of a statistician and are asked to solve problems for companies or organizations. The capstone experiences are designed to make connections among research and statistical topics. Students make decisions based on statistical and mathematical results, prepare a

written report that summarizes the analysis and makes recommendations to the company or organization, and prepare a 5-minute oral presentation that addresses items in the written report.

For post-residency physicians, Samsa and Oddone (1994) organized a statistics course into modules. The first class in each module uses a lecture format and describes the elements that should be included in students' statistical presentations of the topic (e.g., confidence intervals, test statistic, and p-value). The second class uses a discussion format in which students present data analyses and journal article critiques. A series of guidelines are also available for reporting statistical results in medical journal articles (Bailar & Mosteller, 1992). One guideline says to provide enough detail for the statistical methods so that if someone had access to the data they could replicate the study's results.

At an even broader level, writing was incorporated into an entire teacher training program for all PhD students in an arts and science graduate school (Hertzberg, Clark, & Brogan, 2000). The intent of the program is to prepare future statistics teachers. Students receive formal instruction in pedagogy through two courses. The first course uses writing as a pedagogical tool, and a second course focuses on pedagogy in the students' specific discipline area (e.g., biostatistics). Later in the program, students become teaching associates and practice their communication skills in statistical consultation settings.

The above illustrations are from statistics courses or programs in which students have strong mathematics or statistics backgrounds. Instructors of statistics courses for undergraduate or graduate students in social sciences and education face a different situation when they try to incorporate writing into their classes. These students do not usually have strong math backgrounds. They often lack confidence in their quantitative abilities, view statistics as irrelevant to their career, and have a high degree of anxiety towards statistics and research (Onwuegbuzie, 1997). Stromberg and Ramanathan (1996) identify five reasons why students in introductory statistics classes have difficulty when writing about statistics: 1) lack of understanding of material, 2) unfamiliar with technical writing, 3) unable to develop cogent arguments from facts, 4) instructions not followed, and 5) multiple drafts not written. In order to improve students' writing ability, they incorporate journal writing and use note cards with one-sentence answers to various questions. They also ask students to find a survey in a newspaper or magazine and write a letter to the editor discussing the positive and/or negative aspects of the survey. Students evaluate each other's first drafts in terms of the organization, sentence structure, word choice, mechanics and grammar. Each component is graded as 0 or 1. Data showed that the final grades for this project improved when peer evaluations were used.

Another example from an introductory statistics course describes several projects organized around one set of data (Holcomb and Ruffer, 2000). For each project, students work in groups to analyze data and respond to a series of questions. A 100-point analytic rubric is used to evaluate the groups' projects. Mathematical and statistical accuracy accounts for 50 points, and the quality of writing (clarity, organizational layout, mechanics, thoroughness, spelling/punctuation, and professionalism/style) accounts for the other 50 points. The rubric is shared with students so that they know what is expected of them.

Finally, two excellent books were recently published on writing about statistics (Miller, 2004, 2005). As Miller explains, they are not intended to be statistics textbooks nor writing manuals, but instead they describe how to communicate numerical results so that readers can understand how they answer a question. "A naked number sitting alone and uninterpreted is unlikely to accomplish its purpose." (p. 5) The first section of the 2004 book presents a series of principles for writing about numbers, such as setting the context, stating statistical significance, using examples, and defining terminology. The second section contains suggestions on the numbers to report, use of conventions and standards, creation of tables and graphs, and incorporation of good examples. The third section of the book focuses on writing reports for scientific and non-scientific audiences. The second book, organized in the same

manner, is about writing in advanced statistical topics (Miller, 2005).

1.2 The Value of Incorporating Communication into Graduate Statistics Classes in Education

The purpose of this article is to describe an approach to teaching statistics that emphasizes communication and reasoning. The course is for graduate students in education who are teachers, school administrators, counselors, or school psychologists. These students are frequently asked to write papers throughout their coursework, but they have limited experience in writing about quantitative results. A typical student's first impression of a statistics course is that it primarily involves computational and procedural knowledge. A commonly held belief by students is that once a correct numerical answer is obtained or after a correct statistical decision is made (such as rejecting or not rejecting a null hypothesis), then their work is finished. Moreover, when students are drafting their dissertation chapters, they may be quite eloquent in their description of the study and in synthesizing related literature, but they struggle with how to explain the statistical procedures they used to answer the research questions and how to communicate their interpretations of results to others.

For nearly two decades, the mathematics education community has recognized that communication is essential to the learning, understanding, and doing of mathematics. In fact, communication is one of the standards described in publications of the National Council of Teachers of Mathematics (NCTM, 1989; 2000). In the *Principles and Standards for School Mathematics* (2000), the Communication standard states that instructional programs should enable students to: "organize and consolidate their mathematical thinking through communication; communicate their mathematical thinking coherently and clearly to peers, teachers, and others; analyze and evaluate the mathematical thinking and strategies of others; and use the language of mathematics to express mathematical ideas precisely" (NCTM, 2000, p. 60).

Although the NCTM documents were developed for K through 12 educators, the benefits of incorporating verbal, written, and graphical forms of communication into instruction and assessment are the same for educators at all levels. Most importantly, it forces students to go beyond the numerical answers and focus on the *meaning* of the results. Once students begin to verbalize their thoughts and reasonings, the misconceptions and confusions are brought to light and can be addressed by the instructor. With frequent opportunities to write and talk about statistics, students' conceptual understanding is ultimately improved.

When students share their interpretations of data analysis with each other, they realize there are multiple ways to describe and present results. Each student's interpretation is different. There are inaccurate ways to describe results, but there is no one correct way. Radke-Sharpe (1991) describes this benefit as creativity. Reading published research also allows students to see the many different ways of displaying and writing about results.

Learning how to communicate will help educational practitioners in their future careers as they are called upon to interpret results and share them with a variety of audiences in their school districts, including parents, teachers, counselors, administrators, and community leaders. Numerous quantitative reports come across their desks, and the ability to understand what is in those reports, evaluate the accuracy of results and conclusions, and communicate them to others is an essential skill.

Finally, by fostering a classroom environment that encourages students to express their statistical thinking, reasoning, and interpretations in a variety of forms and contexts, students' confidence levels are increased and anxiety is reduced. A large literature base has formed over the years surrounding

statistical anxiety (e.g., Bell, 2001; Onwuegbuzie, DaRos, & Ryan, 1997; Zeidner, 1991). As an example, Onwuegbuzie (1997) delineated four dimensions of statistical anxiety that graduate students may experience: 1) perceived lack of relevance of statistics to their future careers, 2) fear of statistical language, 3) fear of application of statistics knowledge, and 4) interpersonal anxiety, which is the fear students experience when they have to ask for help from peers or the instructor. In a class in which the norm is to communicate statistically by speaking, writing, reading, and listening, students' fears may be eased, and thus their confidence in their ability is improved.

A mathematics educator made the following reflection when she was incorporating mathematical discourse into her lessons.

> "When I started asking students to explain themselves more often, they were uncomfortable. They hesitated, sometimes so long that I did not know whether to keep waiting patiently or to let them off the hook. I remember many times telling them, 'I'm not saying you are wrong, we're just interested in your thinking,' but this statement did not seem to reassure them very much. Quite a bit of time and perseverance were needed to get the students to accept this habit (Van Zoest & Enyart, 1998, p. 153)."

Fortunately, the outcomes are usually worth the initial time and effort. As the teacher further reflects,

> "When they realized that explaining was not frightening after all, so many students wanted to explain…that we could barely accommodate all the sharing they wanted to do. ...When we spent more time discussing thinking, an added bonus was that students were more willing to let me know when they could not understand something. The stigma of 'not getting it' was lessened (Van Zoest & Enyart, 1998, p. 153)."

A middle school math teacher made this reflection about the students in her classroom. College and university educators who are encouraging communication in their statistics classrooms may have similar sentiments.

1.3 Overview of Paper

The first purpose of this paper is to describe a systematic, integrated approach to incorporating statistical communication throughout an intermediate statistics course in graduate education. Students in the course are in the educational administration, instructional leadership, counseling education, or school psychology doctoral programs. The majority has limited mathematics and statistics backgrounds. Most have full- or part-time employment as teachers, principals, counselors, or school psychologists. In their future careers as practitioners who must interpret and communicate results or as researchers who conduct their own studies, it is essential that these students have strong statistical reasoning skills, solid conceptual understandings, and the ability to make accurate and complete interpretations of results.

The instructional approach was designed with the overarching goal to improve students' verbal and written communication in the language of statistics. Two additional goals were to enhance students' statistical reasoning and conceptual understanding and to increase their level of confidence in doing statistics and talking about it with others. The framework for developing the approach was guided in part by NCTM's Communication standard and the description of what students should be able to do (2000).

This paper adds to the body of literature on statistical writing in several ways. The approach is broad. Students are communicating verbally and in writing throughout the entire course, not just on one or two

papers or projects. Detailed illustrations are provided to describe what the instructor is doing, what the students are doing, and the intent is to paint a picture of the classroom environment. For each statistical topic, students see a variety of written examples that interpret results. Examples are from previous students or from published research articles. They vary in their quality and use of text, tables, graphs, and other forms of representation. Students also receive a great deal of feedback from the instructor and from their peers. The instructor gives individual feedback on written communication. Students are able to practice their verbal communication skills weekly through whole-class discussions and small-group activities with their peers. After students' become more familiar with communicating statistically, they participate in a scoring activity in which they evaluate sample paragraphs of results. They score each paragraph as low, medium, or high quality based on rubric criteria they develop. For each paragraph, they provide a rationale for why they assigned the particular score. Finally, to show evidence of the successfulness of the approach, quantitative and qualitative data is provided.

2. Instructional Approach

Four areas describe how the approach integrates statistical communication throughout the course: 1) weekly individual assignments, 2) in-class, small-group activities, 3) use of published journal articles, and 4) a "scoring" activity where students develop rubrics and evaluate paragraphs of results. Detailed illustrations are given within each area, along with the benefits they provide to students and teachers.

2.1 Individual Assignments

Instruction focuses on the importance of including statistical evidence to support statements made about results, writing conclusions that refer back to the original research questions, and communicating results in a way that they can be easily understood. Weekly assignments give students the opportunity to practice these skills and receive feedback on their communication.

Because most students come to the course not used to writing in a "math" class, the first few assignments are more structured than the later assignments. Early assignments ask them to state the independent and dependent variables, indicate the null hypothesis, summarize the descriptive statistics, state the F-values and degrees of freedom (df), state the statistical decision regarding the null hypothesis, and write a one-sentence conclusion about the results of the study. After they become familiar with writing statistically in their own words, the prompts on the assignment become more open-ended. For example, "write a paragraph or two that describes and interprets your statistical analyses as though you are writing in the results section of a research journal article" or "write a paragraph as though you are explaining the results to someone who has a limited amount of statistical knowledge."

2.1.1 Stating Conclusions

This illustration is from an assignment early in the course that asked students to conduct an analysis of variance (ANOVA) on data from a study examining the effects of violence in television on the aggression of viewers. There were three groups of participants. All groups watched a 30-minute segment of a television program, but each group's segment had a different number of violent scenes. After viewing the segment, participants took an aggression inventory.

Discussion of the completed assignment the following week focuses on the one-sentence conclusions. Students received written, individual feedback on their own conclusions, and now they have an opportunity to see other examples of written work. Using previous students' samples, I organize a wide variety of sentences into categories. We discuss each category, and students can see where their own sentence fits.

One category is labeled as "too general". These statements could apply to any study that shows a statistically significant difference among means (e.g., "At least one pair of groups is significantly different" or "The null hypothesis of equality of means was rejected."). Another category is "incomplete/unclear" (e.g., "The groups showed different reactions in aggression to the independent variable"). This statement does not describe the independent variable. A third category is "incorrect" (e.g., "There is no significant difference between violence and aggression".) This is a common misconception of students who are beginning to learn about inferential statistics. Although they understand that an ANOVA tests for significant differences, they incorrectly believe that it examines the difference between an independent and dependent variable rather than a difference among means for levels of an independent variable.

Examples of "good" statements are shown below. It is helpful to provide students with several in order to demonstrate that there is more than one way to write a complete and correct conclusion. All the statements below specifically mention the independent and dependent variables and correctly state that a significant difference was found. Although all of the statements fail to include the word "mean" when referring to a significant difference, the last statement is the most complete.

There is at least one significant difference in aggression scores among groups who were exposed to different amounts of violence in a movie.

The amount of violence viewed does have a significant effect on the amount of aggression displayed.

There is at least one significant difference in aggression scores across groups that watched a movie with differing amounts of violent scenes.

The number of violent scenes during a 30-minute television segment has a significant effect on the amount of fantasy aggression displayed as shown on the Thematic Apperception Test.

2.1.2 Editing Conclusions to Improve Their Quality

A few weeks into the course, students begin to examine lengthier conclusions. In a homework assignment on post hoc testing, students analyzed tax return data. The dependent variable was the efficiency of processing tax returns, defined as the number of days between receipt of the tax return and final processing. The independent variable was the region where the tax-processing center was located: east, mid-west, south, and west. One question on the assignment asked students to "write a few sentences to describe the results of the Tukey test. After receiving individual feedback on their own paragraphs, students were given several complete and correct paragraphs written by previous students. Two of them are shown below:

Tukey post hoc analysis, using a significance level of .05, revealed one significant different between the centers. The regional tax-processing center in the East took more days (mean=54.9, sd=9.8) to complete final processing of tax returns than the tax-processing center in the West (mean=44.7, sd=7.6). There were no significant differences between the mean length of time required for processing between east and mid-west, east and south, mid-west and south, mid-west and west, and south and west.

Follow-up Tukey analyses indicated that the Western tax-processing centers were significantly more efficient than the Eastern processing centers in terms of days needed to process tax returns (p<.05). The mean numbers of days for the East and West centers were 54.9 (sd=9.8) and 44.7 (7.6), respectively. Results revealed no significance for the remaining five pairwise comparisons of

regions.

Both paragraphs explicitly describe the dependent variable, indicate the direction of the significant difference (East had a higher mean number of days than the West), include the sample means and standard deviations, state the significance level used, and account for all unique pair-wise comparisons tested in the Tukey analyses by stating that the remaining pairs were not significant.

Next, students were given examples of incomplete or incorrect paragraphs and asked to describe how they can be improved. For example, the following paragraph correctly indicates the significant difference between the East and West centers and correctly states the dependent variable. However, it does not indicate which region was more efficient nor mention the other regional comparisons. Students said that a person reading this paragraph might assume that the other pairs were not significantly different, but it should be stated explicitly.

   A post hoc test was performed on the data. Results indicate that there is a significant difference between 2 of the groups in terms of efficiency of tax return processing. The East and West groups were significantly different at the .05 level.

Another paragraph, below, does not mention the dependent variable. Students said a reader would not know what was being compared. Also, they indicated that one of the main problems with the summary is that it only describes the statistically based decision regarding the null hypothesis. One student said, "...it does not state what the rejection and non-rejection of the hypotheses means in terms of the variables in the study." In essence, there was no real conclusion. A correct aspect noted by students, however, was that all pair-wise comparisons were addressed.

   Using the Tukey, there is one significant pair-wise comparison between the East and the West. Therefore, the null hypothesis should be rejected. The results of the study revealed that we failed to reject the null hypothesis for the pair-wise comparisons between East and Midwest, East and South, Midwest and South, Midwest and West, and West and South.

Through the individual assignments and the discussions that follow in class, students become familiar with the elements that constitute a complete and correct interpretation. As they edit the paragraphs to improve their quality, students learn how to be precise when communicating. This process encourages self-assessment and reflection of their own work. A benefit for instructors is that it gives them a more accurate perspective of what students know, and do not know, about a statistical concept. They are made aware of misconceptions that often remain hidden when only a numerical answer or statistical decision is required.

2.1.3 Making and Testing Predictions

The individual assignments are not only used for conducting analyses and writing summaries of results. Students are also presented with cognitively rich questions and problems that require them to think further about the problem situation. For example, in the violence and aggression assignment described in Section 2.1.1, students were asked to predict what would happen if two aggression scores were removed and the analysis was conducted again. To test their prediction, they obtained new output, identified the values that changed, and described why.

Students noticed that the new group means were closer together, and they also recognized that the Mean Square between (MSb) and error (MSe) changed. Most students correctly described that MSb was smaller because there was less variation among group means. However, some students were not sure

why MSe was also smaller. The homework sparked an interesting discussion during the following class session when students shared their reasons about why MSe changed. One student explained that, "there was less variance in aggression scores within each of the two groups that had scores removed." Someone also noted that the degrees of freedom for error and total were smaller. The class talked about how these changes would have a bigger impact on the results of a small data set versus a larger one. A similar conversation took place with regard to the magnitude of the values deleted from the data set.

This question on the assignment encourages students to think more deeply about the relationships between the elements in an ANOVA summary table. By doing so, they develop a conceptual understanding of the F-statistic and the crucial role of variance in inferential statistics. At the beginning of the course, not all students are able to write or verbalize their thoughts. However, given repeated opportunities through weekly assignments and whole-class dialogue about them, students gradually develop the ability to communicate their reasoning and thinking.

2.2 In-class Activities

During each class, students participate in small group activities that give them immediate practice in embedding the newly learned topics into their minds and stimulate statistical discussion. They begin to develop ways of verbally communicating with each other that feel comfortable to them. Students who are less comfortable contributing to a statistical discussion in a whole-class setting especially benefit from the chance to ask questions and get help in a small-group environment of their peers. Written communication skills are also enhanced because the group must decide how to clearly and accurately record, in writing, a summary of their group's discussion. A benefit to the instructor is that it creates time in the class session to interact with students on an individual and small-group level, to give immediate feedback, and to determine students' strengths and weaknesses.

A variety of scenarios and problem situations are used. Some activities require them to compare various data situations and the statistical procedures that can be performed. Other activities allow students to make connections among ideas learned in previous class sessions or make sense of numbers provided in output from statistical software. We talk about how the software rarely gives "errors" and that it is the responsibility of the user to select the appropriate analysis to answer the research question and determine whether the variables have the appropriate characteristics. Below are two illustrations of in-class activities.

2.2.1 Making Decisions about Appropriate Data Analysis

In an activity following instruction on the homogeneity of variance assumption, students compared three data scenarios. Each scenario had three levels of an independent variable, but the sample sizes and standard deviations differed. Students first used the Fmax test to evaluate the assumption for each scenario. Then, they were asked to determine whether the validity of the ANOVA results was in jeopardy for those scenarios in which the assumption was violated. In one scenario, students stated that because the group sample sizes were approximately equal (the ratio for the largest versus smallest group size was 1.12), the F result was robust to violation of the assumption. In the other scenario, students considered the group sizes to be unequal, and because the group with the largest sample variance had the smallest sample size, they said the F-test result would be liberal. This meant that if the null hypothesis for an ANOVA were rejected, there would be a concern about the validity of results because of the possibility of a Type I error.

The final question asked students to describe the possible alternative plans of action if violation of the assumption was of concern. Alternatives suggested by students included using a nonparametric test such

as Kruskal-Wallis, using a parametric ANOVA with a more stringent significance level such as .001, or applying a variance stabilizing transformation such as square root or logarithmic.

2.2.2 Comparing Results Across Several Data Sets

This activity takes place in a computer lab using a data set with age and cholesterol levels for 200 subjects. It is referred to as the "original" data set and serves as a point of reference for subsequent analyses conducted on modified versions of the data. Using technology during instruction is beneficial for several reasons. It allows for exploration and discovery. Because statistical software alleviates the need to perform calculations repeatedly, data values can easily be added, deleted, or modified. Instruction then focuses on having students speculate about the changes that occur in the results and the reasons *why* they occur. The intent is not to encourage students to modify data in order to get *better* results, but rather to see how changes to data affect the overall results. It helps students understand the formulas and the relationships among the elements, thus developing a deeper understanding of the concepts. These activities also encourage meaningful student interaction with peers and with their instructor. As instruction progresses, students tend to pose their own "what if" questions.

First, students used SPSS to conduct a simple linear regression on the "original" data set. Means and standard deviations for each variable, regression results, case-wise diagnostics to identify potential outliers, and a scatter plot with the regression line were produced. Students were also asked to save the standardized and un-standardized residuals to the data file. In pairs, they talked about the information that each number in the output provided. For example, there was a moderate correlation between cholesterol and age ($r = .411$ ($p < .001$). Regression results showed that age was a significant predictor of cholesterol ($F_{(1,198)}= 40.31$, $p<.001$), with 16.9% of the variance in cholesterol accounted for by age. One subject with a standardized residual of 4.051 was identified as an outlier. The actual cholesterol value was 520 compared to the predicted value of 279.

Then students were asked to predict how the values in the output would change if the outlier was removed. They tested their prediction by modifying the "original" data set to remove the outlier. Most students accurately predicted that the correlation and R^2 would increase and that the standard error would be smaller. Results showed that $r=.430$ and $R^2=.185$. Some students were surprised that the results only improved slightly. They talked about why this was the case. One student said that although the standardized residual was above 3.00, it was not "that much above it". She speculated that "a bigger difference in r and R^2 would be found if the outlier had a standardized residual of say 9.00 or 10.00". Another student said that there were 200 cases and only 1 was removed. "It would have a bigger impact if the total sample size was smaller, for example, 50, or if more outliers were found and removed."

Next, students were asked to modify the "original" data set by removing a case with a cholesterol value close to the mean of 283.93. The changes in results were negligible, but during the discussion, one group noticed that their output differed slightly from other groups. They realized that two subjects had the same cholesterol level of 283, but different ages. A student explained that the regression equation produces a different predicted cholesterol value for each age. Based on the difference in the groups' results, students determined that the cholesterol level for the 45-year old subject was more accurately predicted than the 52-year old subject. They double-checked this by examining the residuals and found that the younger subject had a lower residual than the older subject.

At this point, students began posing their own questions and modifying the data set to answer them. For example, a few groups wanted to remove subjects with the highest and lowest cholesterol levels. One group removed subjects with the five highest levels, another group removed subjects with the lowest five levels, and a third group removed all ten of these subjects. When comparing results, they discovered

that all three analyses resulted in smaller r and R^2 values compared to the original results of .411 and .169. A student then said that, "it doesn't necessarily matter if a subject has a high or low cholesterol level compared to the rest of the data set. I think what matters more is how accurate the predicted value is. Let's try removing the ten largest absolute values of the residuals." When they did so, the results showed an improvement in the prediction and amount of variance explained (r=.477 and R^2=.228).

Students became increasingly engaged as the activity progressed. In part, this was due to how the activity unfolded. There was a certain degree of systematic progression. It is helpful to have in mind the direction you want to take the activity and the modifications you want students to make. At the beginning, students were told how to change the data set. However, as they became more involved in the activity, they were given the freedom to decide the avenues they wanted to explore.

This activity was used again in a later class session on multiple regression. The statistical software produces an abundance of measures to examine the regression model. Residuals plots are used to check model assumptions and a variety of measures are used to evaluate outliers and influential data points (e.g., hat elements, Mahalanobis' distance, Cook's distance, DFFITS, DFBETAS). The purpose of the activity was not to encourage students to modify data to get a preconceived fit. Instead, the environment was one of posing and answering questions to gain a more meaningful understanding of evaluating a regression model rather than simply memorizing the available measures.

2.3 Examining Journal Articles

2.3.1 Narrative and Visual Representations of Results

Typically, near the end of instruction on each concept, students are given several excerpts from journal articles that show how results can be represented in narrative paragraphs, tables, graphs, and other visual displays. Students examine each representation to determine what details are intended to be conveyed to readers of the article. Students gain practice interpreting results, and it serves as another method of reinforcing complete and correct ways to communicate. It also provides them with a multitude of tools to use when presenting their own research.

As an example, there are many ways to display the outcomes of post hoc analysis. One common method is called the "triangular table format". It is similar in structure to a correlation matrix, but the rows and columns represent groups of the independent variable. Mean differences and p-values are usually provided in the cells of the table. This format allows for results of all possible pair-wise comparisons to be displayed to the reader in an efficient manner. Another graphical way to display post hoc results is by using vertical bars where each bar is a group and the height of the bar represents the mean (Chollet, Delaplace, Pelayo, & Tourny, 1997).

Post hoc results are also displayed within tables of means and standard deviations by using letters. If the same letter is attached to any two means, then the means are not significantly different (e.g., Weger & Polcar, 2002). Other tables use an underlining method instead of letters, to show nonsignificance (e.g., Gerstein & Duffey, 1992). A line is drawn under means that are not significantly different. Finally, tables might have a column that uses "greater than" and "less than" signs to show which groups are significantly different. This table is especially useful when multiple dependent variables are analyzed (e.g., Miller, Greif, & Smith, 2003).

2.3.2 Journal Articles

Approximately twice during the semester, students read and reflect on a published quantitative article in

a research or practitioner journal. They are familiar with critiquing articles in other education courses, but the focus is not usually on the methodology and results. To organize students' reflections, they are given several prompts. Some relate to research design issues, while others relate to the statistical procedures and analysis. Examples of the latter are: Identify the variables in the study and describe what the author's data file might look like (i.e., how many columns are needed, what is the scale of measurement for each variable). How many statistical tests were conducted (e.g., how many ANOVAs were run)? Evaluate the strengths and weaknesses of the results section. What do you like about the way the author described the results? What didn't you like? Were the authors' conclusions supported from the statistical evidence? Are rival hypothesis plausible and did the author provide them? What would you most like to ask the author about the results if you had the opportunity?

One article that I often use encourages rich discussions around several statistical concepts. Cullerton-Sen and Crick (2005) conducted research to examine teacher, peer, and self-reports of physical and relational peer victimization for fourth grade boys and girls. The first set of analyses examines relationships among the multiple perspectives of victimization. Correlation coefficients between teacher, peer, and child reports of victimization are obtained and Fisher's z statistic is used to determine the level of cross-informant agreement for each type of victimization. The second aspect of the article utilizes one-between and one-within repeated measures ANOVA to examine gender differences in children's experiences of relational and physical forms of victimization. Third, the article utilizes regression to assess the unique information provided by teacher reports of relational victimization in the prediction of adjustment.

This study incorporates many variables and analyses, but the authors do an excellent job of clearly and succinctly presenting the results in narrative form and tables. When students first read the article, they may be overwhelmed, but as we break it down into sections and talk about it over the semester focusing on one aspect at a time, they are better able to understand the analyses. Students have reported that it is helpful to see the concepts applied in a real research study. It also helps them to appreciate the amount of work that goes into producing an article of high quality. By discussing full articles, students not only gain exposure to the ways in which researchers and practitioners summarize and interpret their results for various outlets, but they also gain a better appreciation for how the statistical analysis fits into the broader context of a study.

2.3.3 Student-Guided Discussion of Articles

Once per semester, students select an article that utilizes a statistical concept covered in class. Prior to the class session, each student is required to: 1) make the article available to the other group members electronically through the university library, 2) create a list of discussion questions, and 3) read the articles from other group members. On the day of class, each student is responsible for facilitating his/her own article's group discussion using the questions he/she created.

There are numerous benefits to this activity. Students can choose an article that is of interest to them. The small group format encourages student participation. Students feel a sense of ownership when they are responsible for guiding group discussion. Finally, a class session of two to three hours in length allows sufficient time to have a thorough discussion of each article in the group and compare how the same statistical procedure was applied in different research studies.

2.4 Scoring Activity

In this activity, students create rubrics to score other students' interpretations of results. Typically, it is used about two-thirds into the semester after students have had experience writing their own summaries

of data (as described in Sections 2.1.1 and 2.1.2). As students evaluate the work of others they are crystallizing their ideas about the desirable qualities of good written communication, and it deepens self-reflections on their own writing. The steps in creating and implementing the activity are given below.

Step 1: First, an assignment is developed in which students conduct an analysis on a data set, obtain output, and write paragraphs of results. The activity works well with any statistical analysis, and the data set can be simple or complex. The assignment illustrated here involves an analysis of covariance (ANCOVA). The data is relatively straightforward, consisting of an independent variable (type of instruction: computer-assisted or not computer-assisted), a dependent variable (assessment score after two months), and a covariate (aptitude).

There are three parts to the assignment: 1) conduct a one-factor ANOVA to determine if instruction influences performance and describe results in a paragraph, 2) conduct a one-factor ANCOVA using aptitude as the covariate and describe results in a paragraph, and 3) compare the results from the two analyses and explain why they did, or did not, lead to the same conclusion. The instructions regarding the paragraphs can be general ("write a summary as though you were writing for a journal article") or specific ("your summary should include the sample means that are being compared, the F value, df, p-value, the decision regarding the statistical hypothesis being tested, and the conclusion based on the variables in the study.") depending on students' prior experiences in the course.

Step 2: Prepare a set of 10 to 15 paragraphs written by students from previous classes. The illustration shown here includes 11 paragraphs from part 2 of the assignment (see Appendix A). It is important that the paragraphs vary in terms of numerical accuracy, level of reasoning, quality of interpretive statements, use of terminology, accuracy of conclusions, clarity, completeness, and length.

Step 3: When students bring the completed assignment to class, discuss the results with them. This is to ensure that all students have the correct output and an accurate understanding of the ANOVA and ANCOVA results before they move on to evaluating other students' work.

Step 4: Provide each student with the set of paragraphs. They should individually read each paragraph first. This allows them to write their own thoughts about the quality of the paragraphs before discussing them with their group.

Step 5: Provide each group with a blank scoring sheet and rubric form as shown in Figure 1. First, ask them to sort the paragraphs into three levels based on the overall quality: low, medium, or high. More levels could be used, but if this is students' initial experience in developing criteria, three levels seem to work best. Also, it helps to have each paragraph on a separate piece of paper so that students can physically put them into groups and shuffle them during their conversations. After reaching group consensus on a score for each paragraph, the group records the score on the scoring sheet and writes a rationale for why they assigned that score. They also record the scoring criteria they created for each level.

Step 6: After scoring is complete, tally the group scores for each paragraph. The whole-class discussion can begin by focusing on paragraphs that were assigned the same score by all groups. Students can give their rationales for the score and talk about why the response so clearly fit into one of the levels. Then, the focus can turn to the paragraphs with different scores. This is the point where the discussion becomes quite animated as groups support their score and defend their rubric criteria. (see Section 3.2 for a summary of how this activity played out in the classroom).

Paragraph	Score	Rationale for Score
Tonya		
Mike		
Melissa		

Score Level	Description of Criteria
Low	
Medium	
High	

Figure 1. Example of scoring sheet (top) and rubric form (bottom) used in the scoring activity.

3. Impact on Student Learning and Communication

This portion of the paper examines qualitative and quantitative data collected from doctoral students in school psychology and educational leadership/administration programs in order to describe the impacts of the instructional approach on student learning and determine if the initial goals were realized. The data are organized into four areas of evidence: 1) individual student growth over the semester, 2) impact of the "scoring" activity, 3) comparison of interpretive summaries, and 4) student perceptions and confidence.

3.1 Individual Student Growth

All individual assignments and in-class group activities were collected to examine students' solution strategies, explanations, and interpretations in terms of how they progressed over the semester. In addition, two class sessions were videotaped and transcribed in order to examine dialogue among students during the scoring activity, in-class activities, and whole-class discussions.

Students across the ability continuum appeared to benefit from the instructional approach. The first example is Linda. Her work on the first few individual assignments displayed a lack of conceptual understanding and poor communication skills. In whole-class discussions, she rarely participated except when called upon. In small group work on the activities, she preferred to be the "recorder", rarely making a contribution other than to say, "what should I write?" Over time, there was a noticeable change in Linda's performance and confidence level. Her work showed a better grasp of the underlying meaning of concepts, and although her written interpretations were sometimes unclear and awkward, there were fewer errors. She remained hesitant during whole-class discussions, but participated more in her small group. She began to feel comfortable describing what she did not understand and sought feedback from her peers. As I walked into the room before class one evening, I heard her asking a classmate a high-level statistical question about the way in which the group means are adjusted along the regression line when a covariate is in the model. The fact that she was able to communicate the question in a meaningful way, and felt comfortable asking a peer, demonstrated her growth over the course of the semester.

The second example is Jackie. From the beginning of the course, her assignments showed relatively few errors or misconceptions. Explaining answers in written form appeared natural for her. She participated often in whole-class discussions and was actively engaged in her small group. Over time, her verbal communication during whole-class discussions continued to increase. She often responded to questions from other students, which led to meaningful and interactive dialogue. Jackie demonstrated a sense of confidence in her abilities and a willingness to help others understand. Other students often looked to her for help in reasoning and understanding the concepts. She developed a level of statistical maturity that is gratifying to see as an instructor.

3.2 Scoring Activity

The scoring activity took place during a 2 hour and 40 minute class session and was videotaped. Student-created rubrics, scores for each paragraph, and written rationales for each score were collected from the groups. Individual student reflections about the activity were also obtained.

Student engagement during the scoring activity was extremely high and group discussions were cognitively intense. Each group developed its own unique approach to evaluate the paragraphs for correctness and completeness. Some groups evaluated all numerical values first, and then examined the wording. Other groups used a sentence-by-sentence and word-by-word approach. One group had a lengthy discussion of Rhonda's paragraph (see Appendix A). They eventually came to a consensus that the paragraph was a "medium", and their written rationale on the scoring sheet was that the "word choices were often poor, grammar was not great, thoughts could be broken down and more succinct, and the phrasing was not always correct. But all the numerical test results were correct, except for the degrees of freedom." They methodically talked about each sentence. They agreed that the information in the first sentence was mostly accurate, but the wording was awkward and unclear. A student said, "This sentence is crazy. She's trying to bring too much information into one sentence, she needs to break it down." They made corrections to individual words, such as changing the word "covariance" to "covariate" in the first sentence. The group also did not like the phrasing of the next to last sentence in

Rhonda's paragraph that said, "there were statistically significant differences among groups to begin with." They discussed various edits of that sentence to make it accurate and clear. The following are excerpts from the general criteria this group developed and applied to score the paragraphs.

- Low: incorrect data, important values are omitted, interpretation of data or concepts is confusing, failure to explain or describe variables, incorrect conclusions;
- medium: vague but correct interpretation, either too wordy or needs further elaboration, minor numerical reporting errors, failure to completely articulate variables;
- high: reported all data correctly, accurately interpreted results, clearly articulated results, correct conclusions.

Discussions were most lively when students disagreed with each other about a score and had to defend their position. A few brief comments from the transcribed discussions are:

- How did you give Hank a 'medium', when you gave Lisa a 'medium'? Hank's paragraph was much worse.
- Why did he use the words 'treatment effect'? I think a treatment effect is like treating with a drug or an instruction method. That's not what the study is about.
- You can't say there's a significant *difference* between aptitude and assessment. That's not what you're testing. One is a covariate, the other is a dependent variable. They should report the "r" value to show the *relationship*.
- They didn't label and describe the groups consistently.
- He reports original means instead of the adjusted means based on the covariate.
- These sentences are jumbled and misleading.
- Gosh, it's so wordy that you get lost. That's probably me (laughs). I have a lot of trouble being succinct.
- This is really hard. Now I know what it's like for Dr. Parke when she grades our work and gives feedback.

After groups finished, a class tally was taken to compare scores. Conversations across groups were meaningful and mathematically rich. Students shared their rubric criteria with each other and enthusiastically debated scores for each paragraph. They talked about what constituted a high level paragraph and eventually created a sample "complete and correct" paragraph (see Appendix A).

At the end of the course, students were asked to provide a written reflection on the activity and describe the types of paragraphs that caused the most animated discussion in their group, 73% of students indicated that the 'lows' led to good discussions because of the errors in numerical values and the inappropriate interpretations. Other students said both 'lows' and 'highs' were discussed at length: "The examples that were poor generally led to a discussion of everything that was missing, and the examples that were considered good could be used as models to compare our own work to." A few students said they could have used more than three score levels because it was often difficult to choose between a 'medium' or 'high'. They said, "it wasn't like, 'these paragraphs are all high and they are all the same', There are high 'highs' and low 'highs'."

Finally, students were asked to reflect on the ways in which the activity contributed to their learning experience in the course. All students responded positively. Here are a few examples:

- The scoring activity allowed me to see flaws in my own communication in relation to interpreting statistical results. By observing how others interpreted results, it deepened my awareness of how important it is to understand what is being asked in the research question.

- It helped me become more aware of my own writing style and allowed me to self-evaluate.
- The scoring activity really helped me see what sort of explanation was needed to convey all the necessary information for people to understand what you were saying. When reading the examples, I thought about the types of written output I had been providing and learned how to better communicate my understanding.
- It's beneficial to see the right and wrong ways to put statistical data and results into words. This is definitely a very important aspect of this class.
- We were able to take what we learned and actually apply it. In rating the student responses we needed to determine if the student understood the statistical procedure and understood the results. We also rated how well the student was able to communicate his understanding.
- I enjoyed the activity because it allowed me to examine more concrete examples of student interpretation. It also allowed me to utilize feedback that I had gotten and apply it.
- I benefit greatly from the simple act of having others around to 'talk statistics'. The language becomes more fluent, contributing to an overall understanding.

3.3 Comparison of Interpretive Summaries

An opportunity arose to compare interpretive summaries from students who experienced the instructional approach to those who did not. Data was collected in a Measurement Theory and Practice course in which both groups of students were enrolled after their statistics course. An assignment in the measurement course required students to conduct a reliability analysis on a survey that measured attitudes toward computers. Students were asked to conduct a t-test to determine if two groups of subjects (one group from University A, the other from University B) differed significantly in terms of computer attitude. They were required to interpret the t-test results and write a summary as though they were including it in a results section of an article.

It should be noted that this was not a true experimental comparison due to lack of random selection and assignment. There is also a potential confounding variable of instructor. In other words, the instructional approach cannot be separated from the instructor. However, there were many similarities across the two groups. All students were enrolled in the educational leadership doctoral program, had similar occupations, and had similar mathematics and statistics backgrounds. Both groups received the statistics course in the semester prior to the measurement course. They used the same statistics textbook and followed the same sequence of topics. The structure and length of the course was also identical.

A 5-point holistic rubric was developed to score each summary. Criteria in the rubric took into account the accuracy of numerical results, appropriate use of terminology, and the types of statistical evidence used to support the decision that computer attitudes differed significantly. Several precautions were taken to minimize bias. The instructor and an assistant with a statistics background scored the summaries. Office staff blinded the summaries to remove names and other potentially identifying information from the pages so that the raters would not know the students or the class from which they came. The rubric was developed and agreed upon by the raters. Scoring took place after the measurement course was completed and grades were assigned. Inter-rater reliability was 91%. When disagreements occurred, raters reached a consensus on the scores.

Results showed a significant difference in the mean quality of summaries across the two groups. Students who experienced the instructional approach had a higher mean score M = 4.00, sd = 1.07, n = 28) than students who did not experience the instructional approach (M = 2.67, sd = 1.22, n = 21). The value of the two-sample t-statistic is 4.05 with a P-value of 0.0002. To further examine the difference between the summaries, they were also evaluated using an analytic coding scheme to identify the presence or absence of eight elements. The numbers in Table 1 indicate the percentages of students that

included the element in their summary.

Table 1. Percentages of students who included various elements in their interpretive summary.

Elements of the Summaries	Group experiencing the instructional approach	Group not experiencing the instructional approach
1. mentioned the independent variable (i.e., university)	93%	78%
2. mentioned the dependent variable (i.e., computer attitude)	100%	89%
3. included the level of significance (p or α)	73%	55%
4. included the t-value and the associated degrees of freedom	53%	11%
5. stated that there was a "significant difference" in means	100%	55%
6. provided the two group means and/or indicated that University A had a more positive computer attitude than University B	40%	0%
7. described the null hypothesis and its rejection	53%	67%
8. described the alternative hypothesis and its non-rejection	0%	56%

While the majority of students from both groups mentioned the independent and dependent variables in their summaries (see Elements 1 and 2), results showed that the cohort of students who experienced the instructional approach were more likely to incorporate statistical evidence that supported their statistical decision (see Elements 3 through 6). They also provided the means and standard deviations for the two groups and made a meaningful conclusion. Their summaries interpreted the results in a fashion similar to how it would be described in a journal article.

The cohort who did not experience the innovation tended to focus their summaries more heavily on descriptions of the null and alternative statistical hypotheses (see Elements 7 and 8). The following student's summary does not include statistical evidence, does not provide descriptive statistics for the two groups, and does not meaningfully compare the groups. The null and alternative hypotheses are stated, but interpretive conclusions are not made. For example,

    The significance level is .000. We know that if p is low we reject the null hypothesis that "there is

no significant difference between the computer attitudes of the universities." This would indicate that the alternative hypothesis of "there is a difference in the computer attitudes of the universities" would not be rejected.

3.4 Student Perceptions and Confidence

At the beginning of the semester, students responded to Likert and open-ended survey items about writing expectations in their previous coursework. As expected, the majority of students were frequently asked to write papers throughout their graduate work. However, only a few students said they were asked to write papers that summarized research results. In an open-ended question, 67% of the students specifically mentioned that their weakest area in writing a paper was summarizing and interpreting statistical results.

Near the end of the semester, students responded to a survey about the extent to which the class components helped improve their: a) conceptual understanding of statistics, and b) verbal and written communication in the language of statistics. The percentages in Table 2 indicate the portion of students who chose the options "great extent," or "very great extent."

Table 2. Percentage of students indicating that course components improved their learning.

Course Component	Improvements in Learning	Great or Very Great Extent
Feedback on weekly, individual assignments	(a) improve conceptual understanding	73%
	(b) improve communication skills	82%
In-class, small-group activities	(a) improve conceptual understanding	73%
	(b) improve communication skills	82%
Scoring activity (evaluating sample student paragraphs)	(a) improve conceptual understanding	64%
	(b) improve communication skills	73%

In addition, 100% of the students either agreed or strongly agreed that as a result of their experiences in this statistics course, they feel more confident in *understanding the meaning* of statistical results and communicating *verbally* (e.g., explaining to others, contributing to group discussions). Ninety-one percent of students either agreed or strongly agreed that as a result of their experiences in this statistics course, they feel more confident in their *written* communication of statistics.

4. Conclusion

Learning statistics is like learning a new language. Due to the nature of the subject, statistical information is conveyed in numerical values, symbols, notation, and precise terminology. This paper describes an instructional approach that integrated varied opportunities for learning and communicating across the entire semester. Every aspect of the class required students to think, reason, write, and talk in the language of statistics.

As mentioned earlier, the NCTM *Standards* documents (1989, 2000) create a vision of teaching that places emphasis on problem solving and reasoning, understanding quantitative concepts, and communicating ideas and

results effectively. Although the focus of the *Standards* is on K-12 mathematics, the message they portray is applicable to mathematics and statistics learners of all ages. The Communication Standard describes four important elements of instruction. The first two elements are that instruction should enable students: to organize and consolidate their quantitative thinking through communication, and to communicate their thinking coherently and clearly to others. The in-class activities and individual assignments described in this paper give students the opportunity to reflect on the topics they just learned through verbal group discussions and written explanations and interpretations. The interactions among peers and the instructor within a non-threatening environment help students develop deeper understandings of concepts and increase their level of comfort in talking about statistics.

The last two elements in the Communication Standard are that instruction should enable students: to analyze and evaluate the quantitative thinking and strategies of others, and to use the language of mathematics to express mathematical ideas precisely. By reading and critiquing empirically-based journal articles, students become familiar with how researchers and practitioners communicate results and the conventions they use to incorporate statistical evidence into interpretive statements. Students also gain a better understanding of ways to coherently and correctly use statistical terminology to express themselves. Finally, the scoring activity, which evolved from the author's previous research in middle-school mathematics instruction (Parke, Lane, Silver, & Magone, 2003), provides students an opportunity to apply their communication skills in order to evaluate written paragraphs from their peers. Activities such as these increase students' active participation, help them understand what is expected in a high quality mathematical response, and encourage self-assessment.

To summarize, it is essential that graduate students in education be able to organize and consolidate their statistical reasoning. Instructional approaches that create a climate of discussing, questioning, reflecting, and listening can help to prepare students for their future careers when they are required to communicate statistical information coherently to audiences of colleagues, administrators, parents, media, community, and readers of academic journals.

Appendix A. Set of 11 Paragraphs Evaluated by Students and a Complete and Correct Paragraph

Tonya
Results of ANCOVA showed a statistically significant\difference between adjusted means. Therefore, we reject the null Ho: $\mu_{grade1} = \mu_{grade2}$. There is a statistically significant difference in test score among students who had no computer-assisted instruction vs. students who did not have a computer-assisted instruction when adjusted for the covariate (aptitude). Student who had a computer-assisted instruction did better on a test than those who did not.

Mike
The results show, that after adjusting for the covariate of aptitude, that there were significant effects on assessment scores for both the methods of instruction. The adjusted sample means compared, after accounting for the covariate of aptitude, were no computer instruction ($\bar{X}$=2.686) and computer assisted instruction ($\bar{X}$=4.814). Significant effects were found for the instruction method on assessment scores (F(1,19), p<.05). According to the data, the null hypotheses would be rejected because after adjusting for the covariate of statistical aptitude significant effects were found with relation to assessment scores. In effect, the ANCOVA indicates that aptitude in statistics seems to make a difference in assessment scores regardless of the method of instruction.

Melissa
A statistics course was given with students placed in one of two groups (computer or no computer). After two months of instruction an assessment was given to all 12 students. The mean assessment score for group 1 was 3.5 and for group 2 was 4.0. An ANCOVA using results of an aptitude test as the covariate accounted for in the analysis was run to test the null hypothesis that the mean assessment scores would be equal for each group. Adjusted mean scores were 2.686 grp 1 and 4.814 grp 2. The following values were obtained: F=32.544 at df=1;

p=.008 significant at alpha .05 allowing the null hypothesis to be rejected. Conclusion: The mean assessment scores differ with the method of instruction.

Laverne
Because there is a significant relationship between aptitude and assessment (p=.0001), an ANCOVA was appropriate to perform. The null hypothesis was rejected. There is a significant difference in assessment between the noncomputer section ($\bar{X}$=3.5, sd=1.9) and the computer section ($\bar{X}$=4.0, sd=2.1) when aptitude is a covariate for F (1,9)=11.372, p=.008.

Lisa
An analysis of covariance (ANCOVA) was used to test the null hypothesis that there would be no difference in adjusted mean scores on a statistics assessment for students receiving "no computer-assistance" in a statistics course from the adjusted mean scores of those receiving "computer-assistance", measured at the end of the second month in the course. A measure of statistics aptitude, measured at the beginning of the course, was included as a covariate to control for the effects of existing statistical aptitude. The mean aptitude score for students in the "no-computer assisted" section was 5.667, while the mean aptitude score for students in the "computer assisted" section was 3.667. Adjusted mean assessment scores for the "no computer-assisted" section and the "computer-assisted" section were 2.686 and 4.814 respectively. Results of the ANCOVA suggest that the null hypothesis was rejected. Results showed that there was a statistically significant difference in mean statistics assessment scores, measured at the end of two months in a statistics course, for students in the "no computer-assisted" section and those in the "computer assisted" section when controlling for initial statistics aptitude, F(1,9) = 11.372, p=.008. Thus, the mean statistics assessment scores for students receiving "no computer-assistance" were different than students receiving "computer-assistance" when measured two months into the course, when controlling for the level of initial statistical aptitude for the students."

Paul

After adjusting for prior statistical aptitude, and conducting a one-factor analysis of covariance, the null hypotheses that there is no statistical significance between the computer assisted instruction and non-computer assisted instruction statistics course, in regards to assessment, was rejected F(1,9)=11.372, p=.008. The estimated marginal means for each section of the statistics course are as follows: No computer-assisted instruction = 2.686 and computer-assisted instruction = 4.814. Therefore, with the statistical aptitude being adjusted for, there is a statistical significance in regards to the student's assessment of no computer-assisted instruction and computer-assisted instruction. In fact, when statistical aptitude was adjusted, the mean for computer-assisted instruction rose and no computer-assisted instruction fell in regards to assessment. This would indicate that students preferred the computer-assisted instruction course to no computer-assisted instruction.

Hank
There is a significant relationship between aptitude and assessment scores (F(1,11)=32.544, p=.000). There is a significant difference between two groups (no comp. vs. comp.) on assessment scores. (F(1,11)=11.372, p=.000). As you look at original score means and adjusted mean, the difference between two groups has been significantly changed, which means that aptitude influences the assessment scores in the two types of instruction.

Pete
Sample means on statistical assessment were adjusted using aptitude as a covariate. The adjusted means are 2.686 for no comp. method and 4.814 for computer method of instruction. The decision to reject the null hypothesis is based on obtained F(1,9), 32.544, p=.0001 for the covariate of aptitude and F(1,9), 10.812, p=.008 on independent variable of method of instruction. A significant difference exists between methods of instruction for a statistics class when assessment means are adjusted based on pre-existing aptitude.

Rhonda

One factor analysis of covariance (ANCOVA) with one covariance to determine whether the section of the course (Computer assistance or no computer assistance had an influence on the statistics assessment scores after two

months of instruction adjusting for prior statistical aptitude, was conducted. The adjusted means that were being compared to test the null hypothesis of the population means were as follows: the "no computer" group had an adjusted mean of 2.686 +-0.423 (SE) as when the computer group had an adjusted mean of 4.814 +-0.423 (SE). Data from the SPSS output showed that since the aptitude was significant (F(1,11)=32.544, P=0.0001) it means that there were statistically significant differences among groups to begin with. After controlling these differences (removing them by considering the statistical aptitude as covariance) the results from the SPSS output showed significance differences with regard to assessment score for each one of the groups (F(1,11)=11.372, P=0.008).

Serena
Participants in this study were divided into two groups. Adjusted mean of group 1 was 2.686 and mean of group 2 was 4.814. Group (1,9) = 11.372, p=.008. Reject null, there is a significant relationship between computer assisted instruction and no computer assisted instruction.

James
Results were obtained using an ANCOVA to control for pre-existing statistical aptitude while examining if a group who had received no computer assisted training (M=2.86) and the group that received computer assistance (M=4.814). This test showed a significant treatment effect when controlling for pre-existing aptitude (F(1,9) =11.372, p<.008). We should reject the null, as there is a significant difference between the two groups when controlling for pre-existing aptitude.

A Complete and Correct Paragraph Created and Agreed Upon by Students
A one-factor analysis of covariance with one covariate was used to determine if the type of instruction received (computer-assistance versus no computer-assistance) had an impact on students' assessment scores in statistics after taking into account their quantitative aptitude. All the assumptions of ANCOVA were satisfied including the homogeneity of regression slopes (p>.05), and there was a significant linear relationship between the assessment and aptitude scores (F(1,9)=32.54, p<.001, r=.720). Results of the ANCOVA indicated a statistically significant difference in mean statistics assessment scores, measured at the end of two months, for the two types of instruction (F(1,9)=11.37, p=.008). Therefore, when removing the variance due to quantitative aptitude, the adjusted mean for students receiving "computer-assisted instruction" (M=4.81, SE=.42) was higher than the adjusted mean for students receiving "no computer-assisted instruction" (M=2.69, SE=.42).

Acknowledgments

References

Bailar, J. C. III & Mosteller, F. (Eds.) (1992). *Medical Uses of Statistics, 2nd Edition*, Boston, MA: NEJM Books.

Bell, J. A. (2001). "Length of course and levels of statistics anxiety," *Education,* 121(4).

Chance, B. L. (2002). "Components of statistical thinking and implications for instruction and assessment," *Journal of Statistics Education* [Online], 10(3), (http://www.amstat.org/publications/jse/v10n3/chance.html)

Chervaney, N., Collier, R., Fienberg, S., Johnson, P., & Neter, J. (1977). "A framework for the development of measurement instruments for evaluating the introductory statistics course," *The American Statistician, 31*, 17-23.

Chollet, D., Delaplace, C., Pelayo, P., & Tourny, C. (1997). "Stroking characteristic variations in the 100-M freestyle for male swimmers of differing skill," *Perceptual and Motor Skills, 85*.

Cullerton-Sen, C. & Crick. N. R. (2005). "Understanding the effects of physical and relational victimization: The utility of multiple perspectives in predicting social-emotional adjustment," *School Psychology Review, 34*(2), 147-160.

delMas, R. C. (2002). "Statistical literacy, reasoning, and learning: A commentary." *Journal of Statistics Education* [Online], 10(3), (http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html)

Garfield, J. (2002). "The challenge of developing statistical reasoning," *Journal of Statistics Education* [Online], 10(3), (http://www.amstat.org/publications/jse/v10n3/garfield.html)

Gerstein, L. H. & Duffey, K. (1992). "Workers with AIDS: Hypothetical supervisors' willingness to make EAP referrals," *Journal of Employment Counseling, 29*(3).

Hertzberg, V. S., Clark, W. S., & Brogan, D. J. (2000). "Developing pedagogical and communications skills in graduate students: The Emory University biostatistics TATTO program," *Journal of Statistics Education, 8*(3).

Holcomb, J. P. & Ruffer, R. L. (2000). "Using a term-long project sequence in introductory statistics," *The American Statistician, 54*(1), p. 49-53.

Iversen, G. R. (1991). "Writing papers in a statistics course," In *Proceedings of the Section on Statistical Education*, p. 29-32. Washington, DC: American Statistical Association.

Miller, J. E. (2004). *The Chicago Guide to Writing about Numbers*, Chicago, IL: The University of Chicago Press.

Miller, J. E. (2005). *The Chicago Guide to Writing about Multivariate Analysis*, Chicago, IL: The University of Chicago Press.

Miller, M. W., Greif, J. L., & Smith, A. A. (2003). "Multidimensional personality questionnaire profiles of veterans with traumatic combat exposure: Externalizing and internalizing subtypes," *Psychological Assessment, 15*(2).

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Onwuegbuzie, A. J. (1997). "Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety, *Library & Information Science Research, 19*(1), 5-33.

Onwuegbuzie, A. J., DaRos, D., & Ryan, J. (1997). "The components of statistics anxiety: A phenomenological study," *Focus on Learning Problems in Mathematics, 19*(4), 11-35.

Parke, C. S., Lane, S., Silver, E. A., & Magone, M. E. (2003). *Using Assessment to Improve Middle-Grades Mathematics Teaching and Learning*. Reston, VA: National Council of Teachers of Mathematics.

Peterson, R. S. (1988). "Writing across the statistics curriculum," In *Proceedings of the Section on Statistical Education*, p. 190-91. Washington, DC: American Statistical Association.

Radke-Sharpe, N. (1991). "Writing as a component of statistics education," *The American Statistician, 45*(4). p. 292-93.

Rumsey, D. J. (2002). "Statistical literacy as a goal for introductory statistics courses," *Journal of Statistics Education* [Online], 10(3), (http://www.amstat.org/publications/jse/v10n3/rumsey2.html).

Samsa, G. & Oddone, E. Z. (1994). "Integrating scientific writing into a statistics curriculum: A course in statistically based scientific writing," *The American Statistician, 48*(2), p. 117-19.

Spurrier, J. D. (1997). "A capstone course in statistics," In *Proceedings of the Section on Statistical Education*, p. 40-43. Washington, DC: American Statistical Association.

Spurrier, J. D. (2000). *The Practice of Statistics: Putting the Pieces Together*, Pacific GroveCA: Duxbury/Thomson Learning.

Stromberg, A. J. & Ramanathan, S. (1996). "Easy implementation of writing in introductory statistics courses," *The American Statistician, 50*(2), p. 159-163.

Van Zoest, L. R. & Enyart, A. (1998). "Discourse, of course: Encouraging genuine mathematical conversations," *Mathematics Teaching in the Middle School,* 4(3).

Weger, H. & Polcar, L. E. (2002). "Attachment style and person-centered comforting," *Western Journal of Communication, 66*(1).

Zeidner, M. (1991). "Statistics and mathematics anxiety in social science students – some interesting parallels," *British Journal of Educational Psychology,* 61, 319-328.

Carol S. Parke
Duquesne University
600 Forbes Avenue
412D Canevine Hall
Pittsburgh, PA 15282
Parke@duq.edu