STATISTICS ASSESSMENT IN MULTIDISCIPLINARY CONTEXTS

Susanne P. Lajoie and Andrew Chiarella
McGill University
Canada

As statistical education evolves as a discipline more research involving the examination of statistical reasoning across disciplines is anticipated. For example, statistical investigations can cross into areas of scientific reasoning quite easily. In both situations, research questions are posed; data are collected, analyzed, graphed and interpreted. Instead of integrating statistics in the curriculum there is still a division of labour, whereby math educators are responsible for the teaching of statistics, and science teachers the teaching of scientific inquiry. Cross-disciplinary relationships need to be further examined in terms of our definitions of statistical reasoning and how we assess learning and problem-solving across disciplines. Two case studies will be contrasted to reveal the differences between statistical reasoning in a middle school science classroom and a mathematics classroom.

MIDDLE SCHOOL SCIENCE AND STATISTICS

The case study reported here is taken from a larger study of statistical reasoning in middle school (grade 8) physical science classrooms. In the original study, five groups of three students each were observed in two intact classrooms. Observations consisted of written responses to questions about their interpretations of their data from three separate experiments. For the purpose of this paper the responses from one group from the first experiment are reported. This group was chosen because their responses to the data analysis tasks highlight the difference between a computational approach and a modelling approach. In the computational approach calculations are performed but *not* in order to locate patterns in a set of data. For example, the difference between two data points might be calculated. In a modelling approach, any operations on a data set, such as graphing and calculating, are performed in order to determine if there is an underlying pattern. Here the emphasis is on finding relationships between variables (Lehrer & Schauble, 2000).

THE EXPERIMENT

The students conducted a science experiment on heat energy and its properties. This experiment was a genuine discovery activity because the students had not been introduced to the concept of heat energy and its properties prior to this. As the activity took place in a science classroom the goal was to allow the students to personally observe how different containers insulate heat energy to varying degrees; some insulate well others do not. However, science is *not* simply the "passive process of observing and recording events" (Penner, 2000-1, p.1). As such, statistics can add to the active exploration and interpretation of science data. The study examined *the process of the students' statistical* reasoning as they tried to make sense of a natural phenomenon.

In the experiment, the students placed 100ml of hot water into 4 containers: a styrofoam cup, a softdrink can, a glass beaker, and a calorimeter. The temperature of the water was then recorded every minute for 15 minutes in a table. The students then generated a Temperature by Time graph, essentially a scatterplot, (see Figure 1 for an example) with the data for all four containers on it. Given the experiment, statistical goals could have included the correct translation of the tabled data into a scatterplot, the addition of trend lines to the scatterplot, and the interpretation of the trend lines. However, there were no explicit statistical goals in the original activity presented in their textbook.

The data the students collected had two important features with regards to its analysis. First, the relationship between temperature and time for each container was curvilinear. Second, the initial temperatures were *not* the same for the different containers.

PATTERNS, IDEALS AND MODELS

The integration of science and statistics in educational activities such as the experiment described necessitates several types of student assessment. For instance, one could assess the students' understanding of the natural phenomenon studied, their ability to conduct the experiment efficiently, their statistical calculations and graphing, and their statistical reasoning process, including their ability to make predictions about the phenomenon studied. The research presented focused on the students' statistical reasoning. To assess statistical reasoning, the students were asked to describe the pattern in the data, explain how they determined the reported pattern, and provide a measure of their confidence (Likert scale, 1 to 5, where 5 is "total confidence"). Once students had completed the experiment and constructed their graphs they were asked the following questions pertaining to the relationship between temperature and time, and the difference between containers (Styrofoam cup and glass beaker). After responding to these questions the students were also asked to predict the temperature of 100ml of water after 5 and 6.5 minutes had passed. Students were provided with the initial temperature of the water and the type of container. The researcher then conducted the "experiment" in the classroom in front of the students. The students then compared their predictions to the actual outcomes.

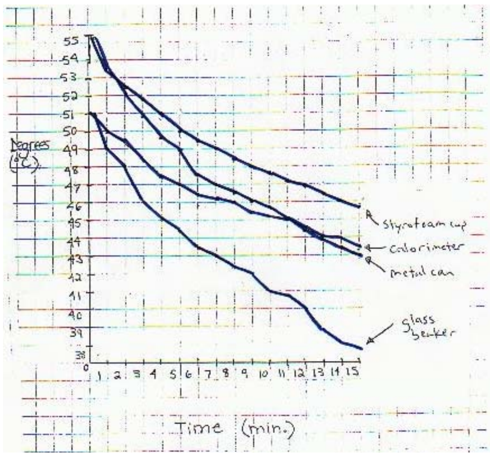

Figure 1. Student Example of a Temperature by Time Graph for Four Containers.

The students also responded to two forced choice questions. The students were presented with four graphs of the same artificial data for the same experiment they had performed. Each graph was "analysed" differently to determine the relationship between temperature and time: points on the graph were connected with straight lines, the addition of a line of best fit, the calculation of the difference between the initial and final temperature, and the calculation of the slope based on the initial and final temperature. The students were asked to select the method of analysis they believed was the best.

CODING

The pattern that students reported was coded along two dimensions: abstraction and specificity (Table 1 presents only codes for levels found in this data set). Abstraction is the representation of a set of data using something other than the data points themselves. Abstraction therefore requires that some kind of aggregate quantity or quality be described that reflects the data set as a whole. Because the coding scheme is a continuum it includes elements that by the definition provided above are not actually abstractions in themselves. They are included in the coding scheme to allow for the description of most student responses. The lowest levels of abstraction in the coding continuum, single data points (level 1) and simple differences (level 2), are hardly abstractions whatsoever. In the statistics education literature the first two levels of abstraction are a result of simply "looking at" and "looking between" the data (Ben-Zvi, 2000), respectively. These levels of abstraction do not go beyond the actual data collected and are therefore not useful for describing the overall relationship between variables like temperature and time, interpolating, extrapolating, or making predictions. Truly abstract patterns begin at the third level of abstraction, rate of change (e.g. temperature decreases 1°C *per* minute). It was expected that the students would not fit abstract patterns but rather use more local patterns. This expectation would be consistent with the first two levels of the coding scheme described. Prior research has shown that students often tend to use specific data points rather than aggregate measures or trend lines (Ben-Zvi, 2000; Hancock, Kaput & Goldsmith, 1992; Krajcik, Blumfeld, Marx, Bass, Fredricks & Soloway, 1998). A possible explanation comes from the work by Kuhn and her colleagues (Kuhn, 1991; Kuhn, Amsel & O'Loughlin, 1988). They concluded that people do not always distinguish between their theories and evidence. In such circumstances data is not seen as possible evidence for or against a theory but rather as a kind of demonstration of the theory. As such, students may be motivated to search the data set for cases (i.e. single data points) that demonstrate their theory rather than evaluate the whole set of data as evidence for or against their theory.

Table 1
Partial Coding Scheme for Reported Pattern: Abstraction and Specificity

Abstraction		Specificity		Description with examples
Changing rate of change	4	Specific	2	Two or more slopes are calculated at different points and compared – “the rate of decrease was 2°C/minute at the beginning and 0.5°C/minute at the end.”
		General	1	Describes different/changing slopes (or rates of change) – “the temperature goes down fast at first and more slowly later.”
Rate of change	3	Specific	2	Calculates change of one variable with respect to the other (i.e. slope) – “the rate of decrease was 1°C per minute.”
		General	1	Describes changes in one variable with respect to the other – “The temperature decreases over time.”
Difference between two data points	2	Specific	2	Calculates difference between two data points – “The temperature dropped 10°C from the start until the end.”
		General	1	Describes the change between two points – “The temperature was lower at the end of the experiment.”
Single data point	1	Specific	2	Reports a data point – “The final temperature was 46°C.”
		General	1	Describes a data point – “The final temperature was low.”

EXPERT PERFORMANCE

The study was not an expert-novice design. The “expert” understanding that follows is based on the researcher's extensive training in statistics and teaching of statistics to others. When analysing bivariate data presented in a scatterplot experts would normally proceed by fitting possible relationships between temperature and time to the actual data. These relationships would be made explicit by the addition of trend lines to the graph or the description of equations with estimated parameters. The stated relationships would then need to be evaluated. For example, the degree to which the data fit a linear relationship might be assessed.

Statistical software was not available, hence in this situation an expert would probably fit a trend line by simply drawing the best line through the data by hand and then assessing how well the “best” line actually matched the data. Given the data in the experiment, the trend lines for all four containers would be curvilinear such that temperature decreases more at first and less as the experiment proceeds (i.e. Time increases). Since the data consisted of temperature and time readings for four containers there would be a matching number of trend lines on the one graph. The expert could then use these lines to describe the relationship between temperature and time for a single container, compare the containers (i.e. which container is the better insulator?), and make predictions.

Obviously, the students in these grade 8 classes were not expected to use advanced techniques or language expected of an expert. For example, students were not expected to know the language of data modelling, such as the different types of general models (i.e. linear, quadratic, logarithmic, etc.) However, it would not be unreasonable to have students draw “best fit” trend lines to the data by hand and then use the trend lines to make decisions. As well, some ability to describe the relationship between temperature and time in everyday terms might be expected. For example, the students could have noticed that the temperature “decreases more quickly at first and then slower later”.

STUDENT PERFORMANCE

The group consistently reported calculating a rate of change (i.e. slope) to answer the questions. They were also confident (a lot of confidence or total confidence) about the slopes they had reported. However, this approach only uses two of the data points (i.e. initial and final) and therefore indicates a poor understanding of error variation. As well, it predicts quite poorly (see below). The group seemed to perform this calculation because they believed this was the correct way to proceed rather than assessing the merits of the calculation with respect to the data. It is not clear if the use of the slope calculation was motivated by the students' belief that calculating is what is expected of them, that they genuinely believed that the strategy was superior, or that they could not think of another way to approach the data. However, when presented with four possible methods in the forced choice questions the group chose the slope method both times. This supports the view that they believed that the slope method was superior in some way. The students may not have even considered the possibility that other, "rival" patterns may have represented the data more accurately. Therefore, they did not attempt to evaluate the pattern they had generated or contrast it with other possible patterns (see Grosslight, Unger, Jay & Smith, 1991, cited in Lehrer & Schauble, 2000).

The group's predictions were quite poor, with errors of -4.8°C and -6.8 °C, respectively. It seems that they had calculated a slope and applied it using the initial temperature and the time interval to predict the outcome temperatures. The students were asked to evaluate their predictions with respect to the actual data results. The group indicated that their predictions were poor. However, *the group maintained total confidence in this method of prediction,* which was the use of a slope. That is, the group seemed base their confidence on the kind of method used rather than its utility to predict. They seemed to be confident because they had used a calculation approach.

Calculating a slope generates a more abstract pattern but results in more prediction error because the relationship between temperature and time is not linear. It may be possible that this group was confident because they had used a somewhat sophisticated mathematical method. Although mathematical, it is not the most appropriate technique given a statistical perspective. This example leads to an interesting instructional question. Should students be encouraged to develop more abstract ways of describing patterns in data given that it may lead to more inaccurate predictions in the short term?

Toth, Klahr and Chen (2000) observed that students' confidence in their conclusions were lower for students who *understood* how to design unconfounded experiments. They speculated that these students' confidence was lower because they had been able to focus on the data and had noted the variation in the results. The present study indicates that confidence may be increased when students believe they are using a sophisticated mathematical approach: calculating for the sake of it versus modelling data with the goal of finding the best fitting pattern or model. Confidence, therefore, may be related to both the perceived sophistication of the method of analysis and the actual goodness of fit between the pattern generated and the data (i.e. magnitude of error variation).

ARE MORE ABSTRACT MODELS BENEFICIAL?

One interesting observation in the present study was that the use of more abstract patterns by the group (and the methods that supported them) did not lead to better results. Specifically, the group presented here, calculated and applied slopes (level 3 abstraction) and generated poorer predictions than other groups who used much simpler patterns (levels 1 and 2). Although this is a concern, the students in the group presented were at least making an effort to engage their data to find overall patterns. This should be encouraged even if it leads to poorer results at times. However, these students did not have a sense of the utility of their patterns for predicting future events. They were not driven to evaluate their pattern or seek out "rival" patterns and potentially revise it given the result of the evaluation.

STATISTICS IN THE MATHEMATICS CLASSROOM

On the surface, statistics in the mathematics curriculum appears to be a natural fit. However, only recently has statistics become part of K-12 instruction (Lajoie, 1998). The following section presents a case study from a grade 8 mathematics classroom. In this previous section we described a science experiment where statistical reasoning could have led to deeper scientific understanding. We selected a case from our mathematics classroom where statistical understanding could lead to better understanding of natural phenomenon.

THE DREAM TEAM CASE STUDY

The Dream Team worked together as a dyad on a tutorial that provided instruction on how to use computer software (i.e., CricketGraph) for representing and interpreting data. The group was familiar with some of the graphs (i.e., column graph, bar graph, pie chart, and line graph) taught in the instruction but not others (i.e., area graph, scatterplot, and area graph). The instructional unit consisted of an exercise on creating graphs that plotted maximum and minimum temperatures daily for the Month of May. Although Dream Team was familiar with the pie chart, scatterplots were new to the group. As such, it was expected that the group might encounter some difficulties. In fact, the group encountered one difficulty: the unfamiliarity of the axes used in the scatterplot for representing May temperatures. The following dialogue reveals that the source of this difficulty was an inability to transfer knowledge acquired in algebra to the current situation. More specifically, it appears that students had problems with the formats of the graphs depicted in each context and recognizing them as the same. For example, axes in a mathematics classroom often appear as in Figure 2a, whereas other types of axes may be used for statistical graphs such as found in Figure 2b.

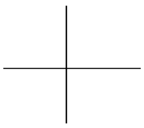

Figure 2a.

Figure 2b.

In the dialogue below, the tutor tries to link the student's prior mathematical knowledge to the scatterplot representation.

Dream 1	Double click on the x-axis. What the h… is that? What's the x-axis?
Dream 2	(calls for tutor) What's the- Help! What's the x-axis?
Tutor	I'll be right back
Dream 1 & 2	What's the x-axis
Tutor	Ok. You guys have algebra yet? Remember, have you ever seen graphs like this in algebra where this is the y-axis?
Dream 1 & 2	Yeah but we're just starting. This is a different variety.
Tutor	Ok this is called a horizontal axis, this is a called a vertical axis. What you have on the screen there is this much of the graph. Ok. See this is the y-axis going up and down,
Dream	Yeah.
Tutor	And the x-axis actually it's here on the positive side. If you were to go down on there that would be the negative y, negative x, but we're working in positive integers so you just go bang bang! So this is the y, this is the x.

This protocol indicates how the teacher links new statistical representations to existing mathematical knowledge. In this situation, the instructional focus is on the graph itself rather than understanding the underlying weather patterns.

INTERPOLATION

The mean in the sample weather graph activity was represented by the "interpolate" function of the scatterplot that plots a line between the minimum and maximum values. The following dialogue illustrates the group's understanding of the mean in terms of the interpolate function.

Tutor	What's that word say- Interpolate
Dream 1	Yeah. Matching the graph
Tutor	See what it did?
Dream	Ohhh!
Tutor	What do you think that was representing?
Dream 1	Look at that
Tutor	For some of the numbers we talked about earlier and some of the statistics. Remember we looked at the median, the minimum, the mean, the range, the maximum

Dream 2	Yeah
Tutor	What do you think that's indicating? Look at the possible-
Dream 2	Like ah to join. Yeah, so you know where ah!
Tutor	I think the minimum would be down here which is at the bottom
Dream 2	You get the mean. The average- like the aver- the high
Tutor	Interpolate means that it's indicating the mean in all of that data.

The scatterplot itself provided a context for constructing relationships since it allowed students to see both the ungrouped (minimum and maximum data points) and grouped (i.e., interpolate line representing the mean) data on one graph. In this sense, graphs such as scatterplots can be used to help students construct meaning of statistical concepts as well as construct relationships among concepts. Construction of such relationships can facilitate the integration and structuring of knowledge and consequently, foster conceptual understanding of statistics.

From an assessment perspective, computers can provide us with a trace of how the group constructed the graphs as well as their final outcomes. However, the graph alone is not sufficient. For this case study verbal as well as written journals were collected to document their understanding. These multiple forms of assessment led us to a better representation of their knowledge. For instance, in their journal, the group acknowledged that the weather data was unusual for the month of May and reasoned that it was "somewhat inaccurate because in May it usually does not get below 5°." In verbal discussions, the group also argued that the data were not typical of May since some of the values were as high as 22° in addition to being below zero. These verbal dialogues confirmed and supplemented information documented in the journal. The verbal protocols indicated that these students recognized that the temperature values were more varied in the month of May than in other months. If this were a science class, a teacher could build on this realization to broaden their understanding of meteorology. Multiple forms of assessment, such as structured journals, verbal dialogues, and computer screen recordings allowed for clarification of reasoning and provides a more complete profile of their understanding than would have been possible using one measure alone.

Two cases were presented from science and mathematics classrooms to demonstrate the usefulness of integrating statistics across disciplines. Instruction and assessment tools will have to be constructed in multidisciplinary contexts in order to document the relationship between scientific and statistical reasoning.

REFERENCES

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *International Journal of Mathematical Thinking and Learning, 2(1&2)*, 127-155.

Hancock, C., Kaput, J., & Goldsmith, L.T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27(3),* 337-364.

Krajcik, J., Blumfeld, P.C., Marx, R.W., Bass, K.M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *The Journal of the Learning Sciences, 7(3&4)*, 313-350.

Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.

Kuhn, D., Amsel, E., & O'Loughlin (1988). *The development of scientific thinking skills*. San Diego: Academic.

Lajoie, S.P. (Ed.). (1998). *Reflections on statistics: Learning, teaching, and assessment in grades K-12,* (pp. 1-336). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., & Schauble, L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology, Vol 5. Educational design and cognitive science* (pp. 101-159). Mahwah, NJ: Erlbaum.

Penner, D.E. (2000-1). Cognition, computers, and synthetic science: Building knowledge and meaning through modeling. *Review of Research in Education, 25*, 1-36.

Toth, E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction, 18(4)*, 423-459.