FACTORS AFFECTING STUDENT DECISION-MAKING ON STATISTICS PROJECTS

Nancy C. Lavigne and Susanne P. Lajoie
University of Pittsburgh, USA, and McGill University, Canada

Statistics has become an integral part of individuals' formal and everyday lives. Experiences that help learners make sense of statistical information are needed so that they can make informed decisions. The view of statistics as a decision-making tool can be emphasized in project-based environments, where students investigate problems that require formulating questions and collecting, analyzing, and representing data to address these questions. Producing investigations in collaboration with peers and presenting results to classmates require that students articulate the understanding that formed the basis of particular design decisions. We found that decisions in this context can be mitigated by factors (e.g., efficiency and social influences) that circumvent the appropriate application of principles (e.g., sampling) in the discipline or practices established in the classroom (e.g., use of criteria to assess peer projects) even though students understand them.

INTRODUCTION

Learning environments that facilitate the development of students' understanding of statistics are needed to help them become knowledgeable consumers of statistical information (Watson, 1997). Statistics is often difficult to learn and understand (Garfield & Ahlgren, 1988; Shaughnessy, 1992). Individuals employ heuristics that, although useful and efficient, can often lead to errors in judgment (e.g., Tversky & Kahneman, 1971). Moreover, they have multiple schemata that are often disconnected and applied differentially depending on the context. A "chance schema," for example, can be applied to a probability problem while a "fairness schema" can be inappropriately applied to an opinion survey problem (Schwartz, Goldman, Vye, Barron, & CTGV, 1998). In this situation individuals reason statistically in one situation but not in another. This occurrence is also seen on problems requiring the same solution (Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993). The computational focus of statistics instruction makes it difficult for students to see the value of statistics in everyday contexts and to make connections between theory and practical applications (Snee, 1993). Students need to learn that statistics is not simply about numbers, but rather deals with numbers in context (Scheaffer, Watkins, & Landwehr, 1998). One way to address this problem is to teach statistics as a practical tool for making sense of the world (Chance, 1997; Snee, 1993). The usefulness of statistics becomes more salient when it is conceived as a process of investigation where (a) questions are posed, (b) data are collected, analyzed, and interpreted, and (c) results are communicated (Lipson & Jones, 1996).

The purpose of this paper is to describe a research project that provides middle-school students with an opportunity to learn descriptive statistics in the context of inquiry. In this situation, students can apply their knowledge of a variety of concepts and procedures to make decisions about an investigation they produce as a group. A number of constraints and factors can play a role in investigations, which can impede students' opportunity to demonstrate the extent of their statistical understanding. Decisions made in the context of investigations can be mitigated by factors that circumvent students' application of disciplinary knowledge. In this paper, we discuss practical constraints and social factors that influenced students' statistical decisions in planning projects. First, we describe the learning environment and then we provide examples of practical and social influences that were observed in our work. Finally, implications for assessment are discussed.

DESCRIPTION OF THE LEARNING CONTEXT

The learning environment that students participated in was part of a larger research project, referred to as the Authentic Statistics Project (ASP). The ASP is intended to help middle-school students learn and think about descriptive statistics in the context of investigation. The investigation process includes generating a research question and collecting, analyzing, interpreting, and representing data to answer the question. Students learn statistical concepts and procedures while engaging in three investigations: a class investigation that provides students with background statistical knowledge and two self-generated group projects that are produced based on students' knowledge. The class investigation modelled the knowledge and skills necessary for carrying out an investigation using descriptive statistics. The class began by discussing ideas for a research question dealing with pulse rates (this activity was expected to engage the students). Students learned about variables, population, and sample within this context. The class then collected the data and compared pulse rates in two conditions, at-rest and after physical activity, learning about sampling, outliers, data analysis measures (i.e., mean, median, mode, range), data organization, graphical representations (i.e., bar, column, pie, scatterplot), and data interpretation. The class investigation took three days to complete. A *library of exemplars* (see Lavigne & Lajoie, 1996) was then shown to reinforce statistical ideas and the investigation process and to provide concrete examples to assist students in producing collaborative investigations. The library (a) describes the goals and criteria (i.e., research question, data collection, data analysis, and data representation) for producing statistical investigations and (b) provides contrasting examples of how previous groups achieved these goals on each criterion. Viewing the library and discussing the examples in class took two days. This activity prepared students for embarking on their two group investigations.

In their own investigations, groups formulated a question that they chose to investigate. These questions (e.g., do music preferences differ for students in grade 7 and grade 10) usually involved surveys (e.g., what is your favorite music?). Groups planned their investigations prior to producing them by addressing prompts on project design worksheets that helped them focus on the relevant content (e.g., what size will your sample be?). Groups were given three to four days to produce each project. Once completed, investigations were presented to the class and feedback was received from peers, the teacher, and a researcher. Some groups produced and presented their projects, one after the other, without an intermediary activity. Other groups were required to evaluate two projects that were displayed on the computer and presented by unknown students before they produced a second investigation. These groups had one day to evaluate computer projects. The purpose of this activity was to provide students with an extended opportunity to apply the assessment criteria and to think critically about the projects in terms of these criteria.

FACTORS THAT ARISE IN INVESTIGATION CONTEXTS

In this paper, we focus on the influences that went into the decision-making processes of six groups while producing their statistical investigations. Three different but related factors influenced these groups' decisions: (a) pragmatic considerations, (b) social influences, and (c) perceived value of feedback. Examples of each of these influences are provided next.

PRAGMATIC CONSIDERATIONS

In statistical investigations, students must be able to make decisions based on their understanding of statistical concepts and procedures and how these apply in particular contexts. We expect decisions based on this understanding to be consistent with disciplinary considerations. In our research, we found that reasons for using particular procedures (e.g., sampling) were frequently based on some knowledge of statistics (M = 58%, SD = 12%). Yet, an equally high percentage of the reasons involved non-statistical considerations (M = 42%, SD = 6%). Some reasons were general (M = 16%, SD = 6%) in that they were superficial in nature (e.g., we used this graph to represent our data in the project because it is better) and linked to previous classroom experience (e.g., we calculated the mean because it was used in class to analyze the pulse rate data). Most of the non-statistical reasons, however, were of a pragmatic nature (M = 26%, SD = 5%). Students' desire for efficiency and awareness of time constraints often took precedence over strict adherence to statistical principles.

Although we prefer that most of our students adopt a statistical approach when designing and engaging in investigations, knowledge of task constraints is an important aspect of problem solving and plays a strong role in the classroom. This point is illustrated in the discussion between Cathy, Jamie, and Alex. Cathy's approach to sampling is more pragmatic and she treats the investigation as a task to be completed for class. In contrast, Jamie is more committed to the appropriate use of statistical procedures; he wants to engage in scientific inquiry and do well. The following discussion ensues when Jamie suggests that the group sample 40 students, ten from four grades (grades 1 and 5 at an elementary school, grades 9 and 11 at the group's high school), for the group's first project. The group intended to compare choice of profession at different age groups.

Jamie: 10 from each grade.
Cathy: 10 for each grade?
Jamie: Why not? The more people, the more accurate!
Alex: Yeah, he's right.
Cathy: And how are you going to put this information on a sheet? 10 from each grade is so many!
Jamie: How is that? It's 40. It's 40 kids. It's nothing! That's so little! We have 45 kids in this grade. It's nothing! Less than 30%
Alex: More is better.
Cathy: Okay, now I have a problem. I can't help. I can't do anything tonight. I can't go to the elementary school. I have to be home for my sister and then I have to cook for my grandmother and get food. I can't, I really can't tonight.
Jamie: But do we have to have it for tomorrow?
Cathy: Yeah, he just said, we have to have the data for tomorrow.
Jamie: Well, then we have to change the thing then.
Cathy: It's impossible. That's what I told you so many times but you wouldn't listen.
Jamie: But we didn't know that it was going to be for tomorrow.

Jamie recognizes the importance of a representative sample but Cathy is concerned about having to go to the elementary school and organizing data from what she considers to be a large sample. Cathy's pragmatic approach ultimately prevailed when the group realized that it would not have time to collect a representative sample. In the end, the group sampled and compared responses from 15 students in grades 7, 9, and 11. Jamie was keen on his original idea and suggested that the group implement it in the second project (i.e., increase the sample size to 40 and include children from the elementary school). This decision would give the group more time to sample from the two schools (i.e., elementary and high school) since they would not need as much time to brainstorm ideas. Cathy resisted saying "No, I'm sticking to grade seven. I'm very sorry but if we are not doing this..." Alex tried to compromise by suggesting "Grade seven and eleven" (i.e., all at one school) to which Cathy responded "No. Grade seven is fine." Cathy's concern for how to practically manage large amounts of data (e.g., how are you going to put this information on a sheet?) continued in the second project. In this case, the group considered whether types of professions should be included in the survey question or whether participants should simply be asked what they would like to be when they grow up. Cathy argued for including categories because "... when you really do the project for like, for a scientific thing, it would [i.e., it is better to leave the question open-ended than to include categories]. And then you could, later you could categorize it right. It would. But now we are doing it in class, and he [teacher] suggested something [possibility of including categories]. And I think that it is a good suggestion and because he suggested it, he can't mark it wrong for anything. It's much easier to ask our question to people whether you have a categorized thing because then we could just cross it off and it would make much more sense."

These examples illustrate that even though students understand that a large sample is more representative than a small sample, practical considerations and students' goals for the task (e.g., engaging in inquiry for learning vs. doing the task for a grade) do not always lead to decisions that are consistent with the discipline. This occurrence is a practical consequence of applying knowledge to a real problem and an aspect of task performance. Perhaps, one strength of abstract problems is that they do not require students to consider practical constraints. Students need opportunities to apply their statistical in both contexts to obtained a balanced understanding.

SOCIAL INFLUENCES

Assessing group investigations requires careful consideration of the nature of student interaction within a group and how it influences group decisions. In our work, the interaction and decision-maker differed across groups. Interaction in two of the six groups was either limited or negative because a single student made the decisions without consulting or taking into consideration the ideas or suggestions of the group. Students in the four productive groups worked well as a team. One group chose to divide the task with students working independently at times, collaboratively at others. Three other groups worked collaboratively at all times; students in two of the groups often disagreed but were committed to achieving group consensus. It is clear from the disciplinary and pragmatic examples that disagreements can reveal much that is ordinarily hidden. Groups can function well as a team, but if explanations are not requested from peers and suggestions are not questioned, the reasoning underlying group decisions remains opaque. In our work, disagreements opened the window into students' views and provided a glimpse of how these perspectives influenced students' reasoning and decision-making. Students can choose to be practical even though they know that their choice is not completely consistent with the discipline. The issue in this case is not a lack of understanding but an awareness of constraints that led to poor judgment. Certain types of group dynamics are more conducive to obtaining accurate information about students' understanding.

Students' understanding of content can also be examined by asking them to evaluate the performance of others based on pre-established criteria. Students can apply assessment criteria in a manner that is consistent with teachers if they are given ample opportunity to apply their understanding of the criteria in various contexts. Lack of anonymity, however, limits the objectivity of evaluations (Mastromatteo, 1993). These two possibilities were borne out in our work. Groups developed a stronger understanding of the criteria each time they produced a project. This finding is based on a comparison of evaluations given to peer projects (i.e., peer evaluation) with those given by the teacher and researcher (i.e., expert evaluation). Table 1 illustrates that with experience evaluations of peer performance became more consistent with those of experts from the first (*M* peer = 19.16, *SD* = 3.05; *M* expert = 17.25, *SD* = 2.27) to second project (*M* peer = 19.85, *SD* = 3.68; *M* expert = 19.13, *SD* = 2.39). Despite this consistency, it appears that anonymity may still have played a role in the evaluations. This possibility is based on group evaluations of projects produced by strangers, which were displayed on the computer (i.e., computer evaluation) and group evaluations of their own performance (i.e., self-evaluation) on projects. Four of the six groups evaluated performances presented on the computer.

Table 1 shows that these groups' evaluations of computer projects were quite low (*M* project 1= 15.7, *SD* = 2.7; *M* project 2 = 10.3, *SD* = 0.8) compared to their peer evaluations (*M* project 1= 19.3, *SD* = 3.17; *M* project 2= 19.5, *SD* = 2.8) and to an expert evaluation of the computer projects (*M* project 1 = 19, *M* project 2 = 14). Groups were more critical in evaluating the work of unknown peers than projects of their classmates. They were reluctant to assign full marks to computer projects even though the criteria were met. For example, one group stated "Very good. They did everything right so we'll take off 0.3 marks." Another group indicated that "the project explained the results very well" but assigned a score of 1.2 out of 2 for that criterion. The stakes involved in evaluating performance in this context were low because the evaluations were never seen by the unknown peers. The situation is different when evaluating classmates who are present and can hold groups accountable for their evaluations. The stakes are even higher when evaluating the performance of one's group due to close collaboration over time. One would expect self-evaluations to be higher than peer evaluations in this case. Such favorable self-evaluations are seen in Table 1 (*M* self 1 = 21, *SD* = 3.49 and *M* self 2 = 23, *SD* = 2.18 compared to *M* peer 1 = 19.16, *SD* = 3.05 and *M* peer 2 = 19.85, *SD* = 3.68) and were found in previous research (Lavigne & Lajoie, 1996). It is possible that groups were not as serious in evaluating their own performance than they were in assessing peers, thereby assigning any score for their own work. Additional research is needed to tease out these possible factors. However, the computer evaluations do suggest that technology can reduce social influences by depersonalizing the task and providing students with more time to review material for evaluation.

Table 1
Mean Evaluation Scores and (Standard Deviations) For Class and Computer Projects

	Project 1 (maximum score: 25)	Project 2 (maximum score: 25)
Peer Evaluation	19.16 (3.05)	19.85 (3.68)
Expert Evaluation	17.25 (2.27)	19.13 (2.39)
Self-Evaluation	21 (3.49)	23 (2.18)
Peer Evaluation (subgroup)	19.3 (3.17)	19.5 (2.8)
Computer Evaluation	15.7 (2.7)	10.3 (0.8)
Expert Computer Evaluation	19	14

PERCEIVED VALUE OF FEEDBACK

The benefits of producing more than one project and receiving evaluations on each are reduced when the feedback received from the evaluations is not valued. Students participating in our research did learn about statistics and improved their test scores (maximum = 51) from the beginning ($M = 13.22$, $SD = 4.17$) to the end of the research ($M = 21.58$, $SD = 5.90$). However, group performance on the projects did not change based on expert and peer evaluations (M difference = 0.99, $SD = 3.18$). Topics for the second projects differed from the first projects for five of the six groups but all involved surveys. The main change groups made from one project to the next consisted of adding a second independent variable (e.g., gender). Other revisions were made as a result of verbal feedback received during the oral presentations of the first project. One group, for example, indicated in its presentation that its sample for the survey "what is your favorite brand name?" was selected randomly. However, the teacher challenged the group to elaborate on its method, revealing that sampling focused on individuals who tended to wear the brand name clothes that were included in the group's survey categories (e.g., Tommy, Nike, Fubu). Classmates suggested that sampling from a class list would be a more appropriate and representative procedure. The group took this recommendation seriously and employed the procedure in its second project. Only this aspect of the group's initial procedures was modified. The group, like all the other groups, did not read the written feedback that accompanied the grades given by the teacher, researcher, and peers even though the teacher stressed that it was critical that they do so since they would be evaluated on whether or not weaknesses identified in the first project were improved upon in the second project. Groups were only concerned with the final score. Only one student thought the written comments would be helpful. She began to read them out loud when a group member told her to look at the total points. She insisted that they read the comments "because that's the way we improve."

In our work, verbal and numerical forms of feedback were preferred over written feedback. The immediacy of feedback provided orally appealed to students' sense of efficiency and practicality. The preference of scores over actual feedback is not surprising given its enhanced value in the educational system. Students' school experience endoctrinated the practice of placing a high value on grades, which can lead to extrinsically motivated students whose engagement on tasks is based on an external reward system (i.e., grades), rather than on internal rewards (i.e., enhanced learning) (Deci, Vallerand, Pelletier, & Ryan, 1991). The preference and effectiveness of verbal feedback provided dynamically suggests that educators need to pay closer attention to this form of assessment in helping students improve while they are learning. The focus on final grades is pervasive and engrained. Systemic changes are needed if we hope to achieve the objective set out by reforms.

DISCUSSION

The middle-school students participating in our research did learn statistical content and procedures and were able to produce investigations based on this knowledge. Our assessment of students' understanding was enhanced when students were committed to explaining their point of view and to achieving consensus. Groups' decision-making in producing investigations was often influenced by pragmatic constraints, while evaluations seemed to be influenced by social factors, such as preserving relationships and an ingrained value for numerical grades. Reform initiatives emphasize forms of assessment that provide rich information about student learning (e.g., NCTM, 1995). Our work suggests that successful implementation of such practices depends, to a certain extent, on context and student conceptions of assessment. If grades are valued above all else, it will be difficult for students engaged in evaluation to articulate the reasons for assigning a particular grade. Part of this difficulty is inexperience with evaluation. Multiple experiences in applying criteria enable students to become more comfortable with them over time. Our work suggests that extended practice facilitates internalization of criteria. However, student perceptions regarding the usefulness of written feedback and their ability to provide such feedback also play a role in how they apply their statistical knowledge in the context of evaluation. Moreover, the social consequences of evaluating known vs. unknown peers may also introduce a subjective bias, regardless of the extent to which the criteria are internalized. In short, assessments must be sensitive to context.

REFERENCES

Chance, B. L. (1997). Experiences with authentic assessment techniques in an introductory statistics course [52 paragraphs]. *Journal of Statistics Education* [On-line serial], *5* (3). Available E-mail: archive@jse.stat.ncsu.edu Message: Send jse/v5n3/chance

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist, 26* (3 & 4), 325-346.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19* (1), 44-63.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24* (5), 392-414.

Lavigne, N.C., & Lajoie, S.P. (1996). Communicating performance standards to students through technology. *Mathematics Teacher, 89* (1), 66-69.

Lipson, K., & Jones, P. (1996). Statistics: Towards the 21st century. In B. Phillips (Ed.), *Statistical education* (pp. 67-73). Swinburne, Australia: International Association for Statistical Education.

Mastromatteo, M. (1993). Assessment of statistical analysis in eighth grade. In N.L. Webb & A.F. Coxford (Eds.), *Assessment in the mathematics classroom, K-12 [1993 Yearbook of the National Council of Teachers of Mathematics]* (pp. 159-166). Reston, VA: National Council of Teachers of Mathematics.

National Council for Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

Scheaffer, R. L., Watkins, A. E., & Landwehr, J. M. (1998). What every high-school graduate should know about statistics. In S.P. Lajoie (Ed.), *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12* (pp. 3-31). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwartz, D.L., Goldman, S.R., Vye, N.J., Barron, B.J., & CTGV. (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12* (pp. 233-273). Mahwah, NJ: Lawrence Erlbaum Associates.

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In D. Grouws (Ed.), *Handbook for research in mathematics teaching and learning* (pp. 465-494). New York: Macmillan.

Snee, R.D. (1993). What's missing in statistical education? *The American Statistician, 47* (2), 149-154.

Tversky, A., & Kahneman, D. (1971). The belief in the law of small numbers. *Psychol. Bull, 76,* 105-110.

Watson, J.M. (1997). Assessing statistical thinking using the media. In J. Gal and J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam, the Netherlands: IOS Press.