Learning Statistics Through Exemplars

Nancy C. Lavigne
McGill University
Educational and Counselling Psychology
Lab of Applied Cognitive Science
3700 McTavish
Montréal, Québec
Canada H3A 1Y2

cz81@musica.mcgill.ca

Susanne P. Lajoie
McGill University
Educational and Counselling Psychology
Lab of Applied Cognitive Science
3700 McTavish
Montréal, Québec
Canada H3A 1Y2

insl@musicb.mcgill.ca

Abstract

This paper implements recent proposals for enhancing the learning of mathematics by developing statistics instruction and assessment for eighth grade students that capitalizes on the use of exemplars. The goal of instruction was for small groups to learn about statistics by engaging in hands-on activities as well as to apply their knowledge and skills by creating statistics projects that involved designing, conducting, and presenting a mini-experiment. Performance criteria which reflected the statistical concepts taught in the instruction were explained to students to ensure their understanding of the task (i.e., project). Groups were assigned to two treatments--exemplars and nonexemplars--which differed in the degree to which criteria modeled the processes of hypothesis generation, data collection, data analysis, and graphic representation. The effectiveness of elaborating on criteria through examples and text (i.e., exemplars) or just text (i.e., nonexemplars) for enhancing learning was examined. Both treatments demonstrated significant performance gains from pretest to posttest. However, students' understanding of representative sampling was significantly better as a result of receiving the exemplars treatment than the nonexemplars treatment. Making criteria more elaborate through examples of performance can thus enhance students' understanding of more abstract statistical concepts such as sampling.

Learning Statistics Through Exemplars

The National Council of Teachers of Mathematics (NCTM, 1989) has proposed that statistics instruction commence as early as elementary school in order to facilitate the development of high-level thinking skills such as problem solving and reasoning. Formal methods of instruction that merely emphasize computational and memorization skills are therefore insufficient (American Statistical Association [ASA], 1991; Mosteller, 1988; Posten, 1981; Shaughnessy, 1992). Alternative forms of instruction and assessment that enable learners to construct their knowledge as well as illustrate and explain their thinking when solving a problem can now be considered. However, the abstract nature of statistical content can pose problems for young learners unless such content is made more concrete and meaningful. Making statistical content less abstract can be accomplished by cognitive apprenticeships (Collins, Brown, & Newman, 1989) that (a) anchor statistical content in concrete examples that model the statistical problem-solving process (i.e., exemplars), (b) guide learners through impasses while they apply knowledge acquired through modeling, and (c) fade assistance when proficiency is attained. The objective of this study was to examine the effectiveness of an instructional method that focused on modeling, one component of the cognitive apprenticeship model, for facilitating secondary students' learning of descriptive statistics. Two research questions were posed: (a) does providing concrete examples of statistical procedures facilitate the learning of abstract content and (b) do students acquire depth or breadth of knowledge.

Theoretical Framework

According to educators and researchers, current statistics education is inadequate due to (a) insufficient conceptual background given to students (Garfield & Ahlgren, 1988; Posten, 1981), (b) an emphasis on the abstract nature of the content (Mosteller, 1988) and (c) a reliance on formal methods of instruction (Posten, 1981). These factors result in (a) a reliance on intuitions or opinions which can cause difficulties in reasoning about sampling (Jacobs, 1993; Schwartz, Goldman, Moore, Zech, Smart, Mayfield-Stewart, Vye, & Barron, 1994; Tversky & Kahneman, 1971) and probability (Kahneman & Tversky, 1973, 1982; Tversky & Kahneman, 1973, 1983) and (b) an understanding of the mean as a computational rather than conceptual act (Pollatsetsek, Lima, & Well, 1981). Such difficulties make developing statistics instruction for grades 5-8 problematic where the proposed content includes measures of central tendency and variation, population, sampling, and anomalies (American Statistical Association [ASA], 1991). One way to address shortcomings of statistics education for enhancing high-level thinking is to provide a learning context in which students are granted opportunities to (a) directly observe expert performance through concrete examples that model statistical problem solving, (b) emulate expert performance by applying statistical knowledge on hands-on activities, (c) focus on interpretation by using computer software for analyzing and representing data, and (d) expand statistical knowledge through prompts presented in the form of questions that encourage further thinking.

Method

Twenty-one eighth grade mathematics students (nine females and twelve males) participated in this study. Students were divided into eight mixed-ability groups, each consisting of two to three students of mixed ability in mathematics. Ability groupings were formed by the experimenters based on the teacher's rating (i.e., high, medium, low) of each student's performance as measured by classroom assessments from the beginning of the year. Each group worked on an Apple®Macintosh™ workstation which was set up in the students' regular mathematics classroom.

A form of cognitive apprenticeship (Collins et al., 1989) was adopted to teach students descriptive statistics (i.e., measures of central tendency and variation, population, sampling, and anomalies). Two phases of apprenticeship were provided. The first phase consisted of modeling procedures, coaching students, and fading assistance on instructional activities that situated learning in worthwhile and engaging problem solving tasks. Two skills were modeled in the activities: statistical problem solving and the use of software applications such as Mystat™ (Systat, 1988) and Cricketgraph™ (Cricket Software, 1989) as tools for computing and representing statistics. Coaching was provided by the mathematics teacher, six graduate students, and prompts which were meant to encourage students to reason about data, to facilitate discussions of statistical concepts, and to extend students' learning beyond the information given (Resnick, 1989; Rosenshine & Meister, 1992). Fading consisted of gradually withdrawing assistance as students attained mastery. The instruction enabled learners to acquire the knowledge of facts and tools required to conduct their own experiment as a group project and consequently to perform on open-ended test essays.

Prior to conducting their own experiments groups of students were randomly assigned to two treatments: nonexemplars and exemplars. These treatments consisted of the second phase of apprenticeship which strictly focused on modeling the process of designing, conducting, and presenting an experiment. More specifically, computer software, HyperCard™ (Claris Corporation, 1991), was used to provide information which modeled hypothesis generation, data collection, data analysis, and data representation. These processes were conveyed as performance criteria in each treatment. The nonexemplars and exemplars treatments differed in the extent to which (a) performance criteria for developing and assessing experiments were made visible to students (Frederiksen & Collins, 1989) and (b) statistical procedures for designing and conducting research were made more concrete. Although each criterion and procedure was described textually in both treatments, only the exemplars treatment provided (a) digitized video clips that modeled hypothesis generation as well as the collection, analysis, interpretation, and representation of data by providing examples of performance of students participating in a similar study the previous year and (b) prompts that guided discussions regarding differences between examples (see Figures 1 and 2). In this sense, the exemplars served as a tool for (a) ensuring that students were aware of and understood the criteria for conducting an experiment and (b) making procedures more concrete through modeling by providing several examples (i.e., digitized video clips) of peers explaining how they designed and conducted their experiment during their presentations. This paper focuses on the effectiveness of the exemplars approach in making statistical procedures less ambiguous through modeling. The effectiveness of this treatment was examined in terms of students' performance on an open-ended pre and post test which was analyzed quantitatively and qualitatively and on projects which was analyzed qualitatively. Since the exemplars approach was intended to situate statistical procedures in concrete examples, it was expected to be a more effective tool than the nonexemplars approach for enhancing statistical learning and engendering the knowledge acquisition of statistical procedures such as hypothesis generation, data collection, and data analysis.

Results

Quantitative analysis of the following were conducted: (a) students' overall test performance to determine whether the exemplars treatment was more effective for enhancing statistical learning than the nonexemplars treatment, (b) students' performance on individual test items to explore whether the exemplars treatment was more effective in fostering knowledge acquisition of particular concepts and procedures than the nonexemplars treatment, and (c) students' performance on individual test items to examine whether knowledge of a few or many concepts and procedures was acquired. Qualitative analysis of written responses to test items and of performance on presentations of projects were conducted to determine whether students acquired depth or breadth of knowledge as a result of the instruction. Finally, inter-rater reliabilities were performed to examine consistency between raters.

Quantitative analysis of students' overall test performance examined whether or not type of treatment (nonexemplars or exemplars) affected students' test scores (pre and post). Results from the Subject{Treatment (2)} x Test (2) ANOVA demonstrated that there were no significant differences between the two treatments ($F(1, 16)=0.010, p>0.05$). However, a significant test effect was found which indicated change in statistical knowledge for all students ($F(1, 16)=50.130, p<0.05$). Students receiving the nonexemplars (Mpre=6.636, Mpost=13.818) and exemplars (Mpre=5.714, Mpost=15.143) treatments acquired a substantial amount of statistical knowledge as a result of instruction.

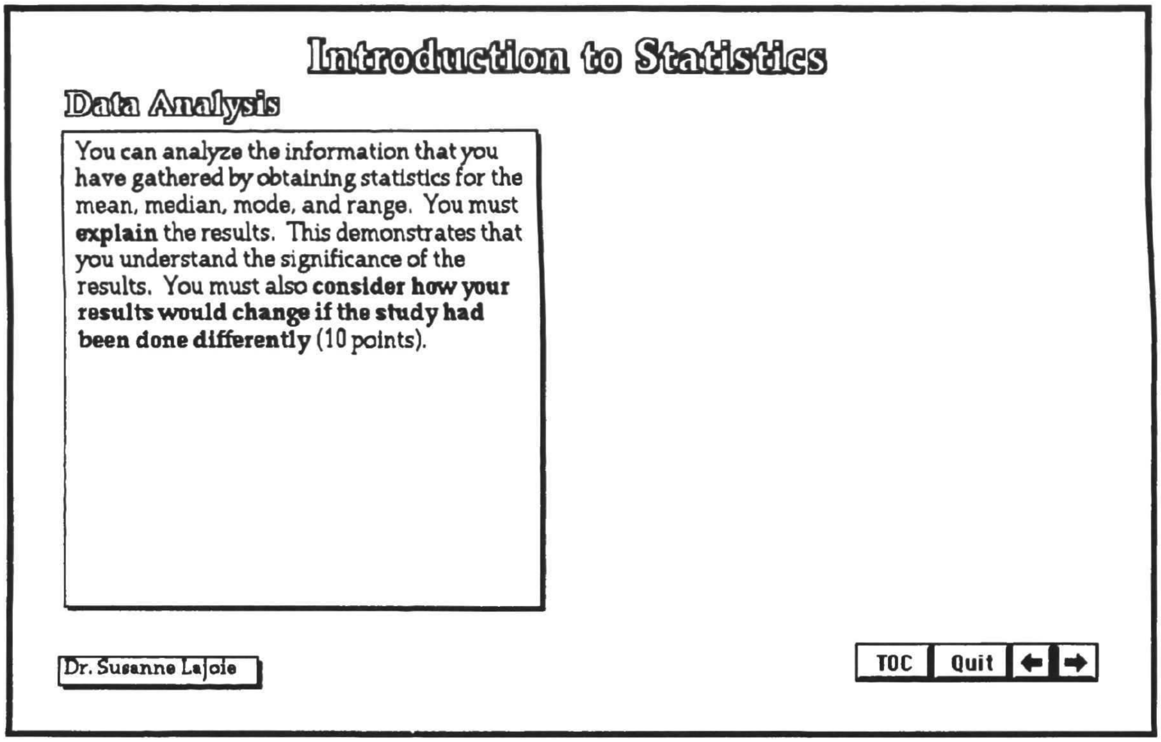

Figure 1: Example criterion provided by the nonexemplars approach: Data analysis.

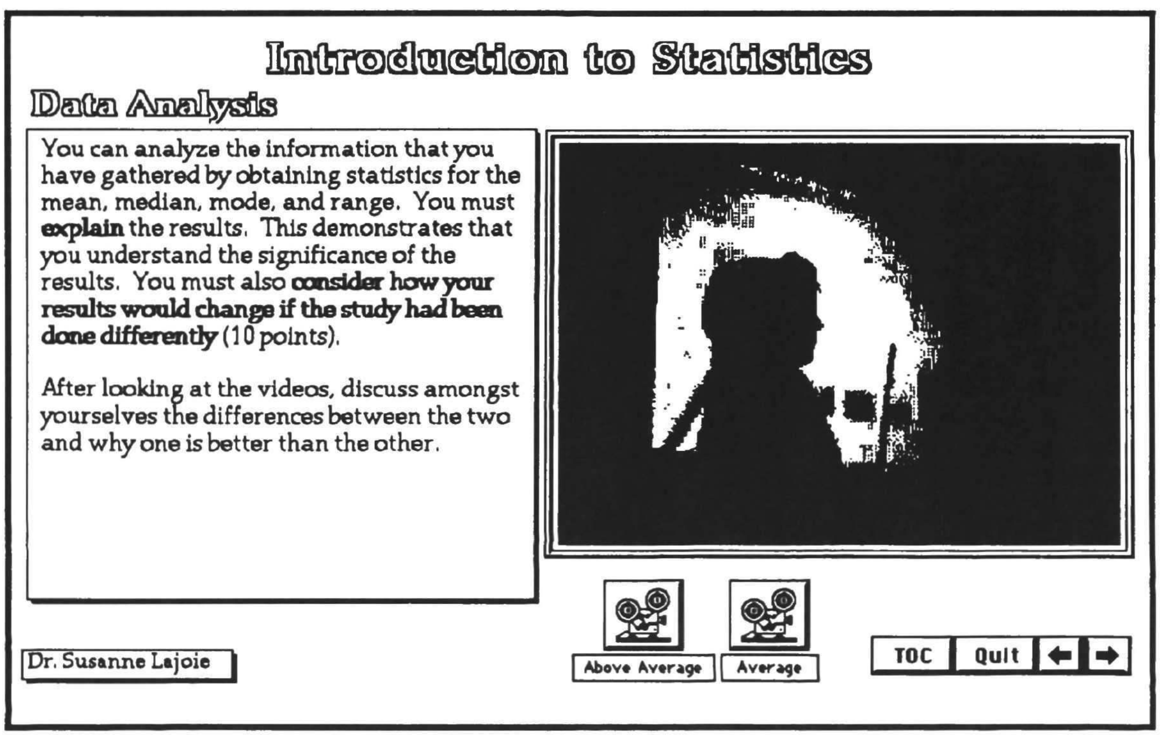

Figure 2: Example criterion provided by the exemplars approach: Data analysis.

To determine whether knowledge gains differed by treatment and whether these were limited to some content or inclusive of all content, individual Subject {Treatment (2)} x Test (2) ANOVAs were performed on test items measuring each of the following statistical concepts and procedures: statistics, data, graph interpretation, outlier, hypothesis generation and identification, population, sample representativeness, sample size, randomization, sample, median, mean, and range. The Bonferroni procedure (Kirk, 1982) was applied to adjust for Type I error in each of these analyses. Significant interaction effects for sample representativeness ($F(1,16)=8.581$, $p<0.01$), sample size ($F(1,16)=6.862$, $p<0.01$), and statistics ($F(1,16)= 4.899$, $p<0.01$) demonstrated that knowledge gains related to sampling differed by treatment. The Scheffé S procedure (Kirk, 1982, pp. 121-122) for making post-hoc comparisons indicated that the difference in performance from pre to post test on items of sample representativeness was significant for the exemplars treatment ($F(1, 16)=24.225$, $p<0.05$) but not for the nonexemplars treatment ($F(1, 16)=2.172$, $p>0.05$). Moreover, this difference was significant at post ($F(1, 16)=6.425$, $p<0.05$) but not at pre ($F(1, 16)=2.581$, $p>0.05$). This finding suggested that students' understanding of representative sampling was facilitated through the use of concrete examples which modeled the procedure. Results for sample size indicated that there was a significant pre and post test difference for both the exemplars ($F(1, 16)=36.413$, $p<0.05$) and nonexemplars ($F(1, 16)=11.388$, $p<0.05$) treatments. Finally, results for the concept of statistics indicated that there was a significant difference between the exemplars and nonexemplars treatments at pre ($F(1, 16)=4.944$, $p<0.05$) but not at post ($F(1, 16)=0.825$, $p>0.05$) and that the difference from pre to post test was significant for the nonexemplars treatment ($F(1, 16)=20.988$, $p<0.05$) but not for the exemplars treatment ($F(1, 16)=0.675$, $p>0.05$).

Significant test effects for statistics ($F(1, 16)= 12.214$, $p<0.01$), graph interpretation ($F(1, 16)=14.473$, $p<0.01$), hypothesis generation and identification ($F(1, 16)=14.141$, $p<0.01$), sample representativeness ($F(1, 16)=22.712$, $p<0.01$), sample size ($F(1, 16)=46.622$, $p<0.01$), sample ($F(1, 16)=12.287$, $p<0.01$), and range ($F(1, 16)=10.670$, $p<0.01$) demonstrated that students acquired knowledge of <u>many</u> concepts as a result of instruction. However, qualitative analysis of written responses to test items (pre and post) and presentations revealed that students' understanding of the content was general which suggested that <u>breadth</u> rather than depth of knowledge was acquired. Not all content was understood by students. Some concepts and procedures, notably measures of central tendency, were problematic for learners. Students were unable to calculate the mean, median, and mode by hand despite knowing how to do so. In addition, students had difficulty distinguishing the mean from the median, often defining the median as the "average" rather than the "middle number," terms which in the instruction, were used exclusively to define the mean and median respectively. This confusion suggested that the concepts mean and median were not fully understood. Moreover, performance on group presentations indicated that most students did not analyze their data. Most groups calculated percentages rather than means for describing their data. Given that all their research questions entailed collecting frequency data this finding is not surprising. However, it is unclear from group presentations whether students understood that the mean was an inappropriate measure for analyzing frequency data. Student responses during the question periods that followed the presentations seemed to suggest that measures of central tendency were not used to analyze the data since graphs were deemed sufficient for conveying the results.

Inter-rater reliabilities were conducted to examine consistency in the ratings given by two graduate students on the pre and post tests. The high correlations for pre ($r=0.982$) and post ($r=0.987$) indicated that the scoring criteria for assessing test performance were clear to raters. This finding suggests that the test results were reliable.

Conclusions and Implications

The present study demonstrates that statistical learning for young learners can be facilitated by a method of instruction that models procedures by providing various exemplars of peers explaining how they engaged in the experimentation process. Substantial knowledge gains from pre to post test were demonstrated. The increase in knowledge is considerable given the four-day duration of the study. Within this time span, students acquired knowledge of many statistical concepts and procedures, however, responses to test items and performance on project presentations indicated that depth of understanding was not acquired. This finding may be accounted for by the high content coverage, limited time in which to learn such content, and emphasis on general skills. According to Pollatsek et al. (1981) and Zawojewski (1988), conceptual difficulties in understanding the mean, for instance, are due to formal methods of instruction which emphasize specific skills such as memorization of algorithms. However, this study suggests that instruction emphasizing interpretation without sufficient experience with computation can lead to difficulties in acquiring conceptual understanding. Although conceptual understanding was not attained, this study suggests that additional modeling through the use of concrete examples (i.e., exemplars) can facilitate students' understanding of representative sampling. Providing students with multiple representations of realistic performance to make abstract concepts such as representative sampling more concrete can therefore enhance learning. However, the choice of representations is crucial. Exemplars must be rich enough to differentiate the levels of performance that are used to illustrate various statistical procedures. Without such differentiation, the effects of making concepts and procedures less abstract will be minimal. This study was limited in that the examples were insufficiently differentiated.

This study was a first attempt at incorporating one aspect of the cognitive apprenticeship method of instruction for eighth graders. However, much more work is required to develop a strong instructional tool that incorporates all the relevant features of the cognitive apprenticeship model. This study is limited by insufficient standardization of the instruction. Scaffolding was provided by graduate students who had to be trained rather than by an instructional medium that provided identical instruction to all students (e.g., computer-based instruction or an intelligent tutoring system [ITS]). Developing such a tool is the next phase of this research.

References

American Statistical Association (1991). *Guidelines for the teaching of statistics K-12 mathematics curriculum.* Landover, MD: Corporate Press.

Claris Corporation. (1991). *Hypercard Version 2.1* [Computer program]. Santa Clara, CA: Apple Computer Inc.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser.* Hillsdale, NJ: Earlbaum.

Cricket Software Inc. (1989). *Cricket Graph Version 1.31* [Computer program]. Valley Stream Parkway, PA: Author.

Fredriksen, J. R. & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18* (9), 27-32.

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.

Jacobs, V. R. (1993). *Stochastics in middle school: An exploration of students' informal knowledge.* Unpublished master's thesis, University of Wisconson, Madison, WI.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80* (4), 237-251.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11* (2), 123-141.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Monterey, CA: Brooks/Cole Publishing Company.

Lavigne, N. C. (1994). *Authentic assessment: A library of exemplars for enhancing statistics performance.* Unpublished master's thesis, McGill University, Montreal, Quebec.

Mosteller, F. (1980). Classroom and platform performance. *The American Statistician, 34* (1), 11-17.

Mosteller, F. (1988). Broadening the scope of statistics and statistical education. *The American Statistician, 42* (2), 93-99.

National Council of Teachers of Mathematics Commission on Standards for School Mathematics (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90* (4), 339-363.

Pollatsek, A., Lima, S., & Well, A. (1981). Concept or computation: Students' misconceptions of the mean. *Educational Studies in Mathematics, 12*, 191-204.

Posten, H. O. (1981). Review of statistical teaching materials for 11-16-year olds. *The American Statistician, 35* (4), 258-259.

Resnick, L. B. (1989). Treating mathematics as an ill-structured discipline. In C. I. Randall & E. A. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 32-60). Reston, VA: National Council of Teachers of Mathematics.

Rosenshine, B., & Meister, C. (1992). The use of scaffolds for teaching higher-level cognitive strategies. *Educational Leadership, 49* (8), 26-33.

Schwartz, D., Goldman, S., Moore, A., Zech, L., Smart, L., Mayfield-Stewart, C., Vye, N., & Barron, L. (1994, April). *Adolescent understanding of sampling in the context of a survey.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

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 Publishing.

Systat Inc. (1988). *Mystat Version 1.0: A personal version of Systat.* [Computer program]. Evanston, IL: Author.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76* (2), 105-110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology , 5* , 207-232.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90* (4), 293-315.

Zawojewski, J. S. (1988). Teaching statistics: Mean, median, and mode. *Arithmetic Teacher, 35* (7), 25-26.