® DATA ANALYSIS OR HOW HIGH SCHOOL STUDENTS "READ" STATISTICS

VERMETTE Sylvain
GATTUSO Linda
Université du Québec à Montréal
Canada

BOURDEAU Marc
École Polytechnique de Montréal
Canada

In most countries, statistics are included in the mathematics curriculum and taught by mathematics teachers. This leads to students learning the elements of statistical concepts as mathematical and to more emphasis placed on being able to compute different measures (e.g. mean, median, standard deviation) rather than their meaning and use. Moreover, in Quebec, the high school curriculum favours a scattered presentation of statistical concepts: tables and simple graphical representations are seen in the first year; averages, medians and histograms in the third; position measures in the fourth and some aspects of correlation and standard deviation are seen in the fifth. Some elements of probability are seen in the second year. But "statistics requires a different kind of thinking" (Cobb & Moore, 1997). Is it possible by making students compute statistical measures to foster the development of statistical thinking and prepare to draw conclusions from different data sets - all important abilities for "reading" statistics, an essential part of communication. This study attempts to determine if high school graduates develop the ability to effectively interpret the use and meaning of statistics (i.e. develop a "statistical way of thinking"). To this purpose we investigated (a) if a mode of data representation (1) list of data, (2) graphical, (3) principal location and dispersion parameters (mean, median, quartiles, standard deviation, etc), influences the students' answers, (b) if students take into account the context of the data in their analysis, and (3) if students' reasoning reveals "statistical thinking" as described in McGatha, Cobb and McClain (1998). A multiplicative argument combined with the use of the context in which the data are presented is preferable to an argument using only one point or only one measure (usually the mean) or only an arithmetic reasoning. To do so, a questionnaire with seven items asking students to choose from two or three samples and to justify the choice was presented to 141 fifth year high school students in the three different modes of presentation mentioned above. The results show that almost one third of the students revealed correct statistical thinking and another 41 % take the whole sample into account. The majority of the explanations are linked to the context. However, for some students more difficult tasks seemed to trigger a more global interpretation.

INTRODUCTION

Communicating statistical results is a "fundamental tool in many careers" (Konold, 2003, p 193) but, in general, understanding and interpreting communications including some statistics is necessary for every citizen. In fact, without this knowledge, it is difficult to have an informed opinion and participate in social and political debates concerning environment, health, education. (Ibid.)

The question is "are our high school students prepared for this task?" Even if they encounter statistics in other areas of study such as geography or economy, for example, most of the time statistics are formally taught in the mathematics classroom. Consequently, statistics are part of the mathematics curriculum and treated as such, emphasize accurate calculation more then fostering the development of the statistical reasoning necessary to communicate or understand information including quantitative data. An analysis of the high school curriculum in the province of Quebec (Canada) reveals that it favours a scattered presentation of statistical concepts (Gattuso, 2001). Even if some statistics are included in the elementary school curriculum, the high school curriculum does not take this fact into account.

In the first year of high school, tables, simple graphical representations and exercises are introduced asking only to read without requiring any interpretation. In the second year, we forget statistics to concentrate on elementary probability. The third year is the year of central measures: average, median, and mode. Continuous data grouped in intervals are introduced and therefore histograms are added to graphical representations. Position measures such as quartiles, centiles and dispersion measures are seen in the 4^{th} year along with box-plots. In the 5^{th} year, standard deviation is added to the dispersion concepts. Some classes may also touch correlation but in a qualitative way by observing scattergrams. In some way, students might acquire all the necessary elements for "exploratory data analysis" as proposed by Tukey (1977) who favoured examining data for unanticipated results over limiting analysis to confirm a priori hypotheses.

Without going into the complexity of inference, it is possible to read and communicate results produced by an intelligent reading of data: this is the basic idea of data analysis. However, in our experience, students tend to see data as numbers without imbedding them in the context, thus neglecting the events they represent. Their school background usually leads students to limit comparison to confronting the means of two data sets. In this study, we investigated first if students, who learned statistical notions separately and in a loosely related manner over their years of high school, develop a statistical perspective (Konold, 1996) and acquire an authentic data analysis point of view using explanations that translate a multiplicative reasoning[1] related to the context of the data set (Cobb, 1999).

A previous exploratory investigation (Vermette, 2004) suggested that the mode of inscription of the data had an influence on the level of justification of the students that serves as an indicator of their statistical thinking. Graphic representation seemed to have a positive effect. We then found it necessary to explore the effect of the mode of inscription on statistical reasoning.

All in all, we wanted to see 1) what level of statistical thinking (as seen through the justification) was reached by senior high school students, 2) if their justifications were related to context, 3) if the mode of presentation had any effect on the level of justification or the use of the context.

BACKGROUND OF THE STUDY

A group of researchers (McLain, 1999; McGatha, Cobb & McClain, 1999; Cobb, 1999) developed and experimentally tested classroom activities focusing on the notion of distribution to introduce the basic notions of central measures and spread-outness as characteristics of distribution using various graphical representations. The results of their experiment show that the class globally evolved from arguments about collections of data points to arguments about distributions. This shift in arguments indicates the development of a data analysis point of view showing the capacity of discussing data sets as distributions describing global differences between two distributions in qualitative terms. The students were in a seventh-grade class.

Moreover, Konold (2003) adds that students need a reason to focus on the characteristics of a distribution, and asking if two samples differ may be a way of triggering more global arguments (Russell, Schifter, & Bastable, 2002). This led us to the construction of a written questionnaire consisting of seven tasks asking students to compare two or three samples in order to make a decision. The context of six of the seven tasks was taken from the study of Cobb *et al*.

The first task, "Batteries" (McClain, 1999), presented data of two different brands of batteries that were tested to see how long they lasted, a context quite familiar to teenagers who

1 We use "multiplicative reasoning" and "additive reasoning" as in Harel, G. & Confrey, J. (1994) and in McLain, 1999; McGatha, Cobb & McClain, 1999; Cobb, 1999. A multiplicative reasoning would use a "proportion" a part of the whole as in " more temperatures are above…" or "90% of the breads …". An additive reasoning, uses quantities and not proportion as in "there are 8 less batteries …"

typically use various electronic devices. The number of data in each sample was 10, a number that could easily lead to a conversion to percentage. The second, “Basketball” presented the number of points scored by two basketball players for eight games and asked students to choose which player should be selected for the all-star tournament (McGatha, Cobb & McClain, 1999). In this case, the number of games may interfere with the use of multiplicative arguments, one eighth (1/8) being more difficult to convert into percentage. The third question, “Trip Decision” (ibid.), provided the temperatures of the city of Boston for a week in September and a week in April for 1994, 1995, 1996. The students had to choose a month for the school trip. The situation for the fourth task, “Breads”, was taken from Raymondaud (1997). We considered this scenario as one of the more complex items because it concerned the production of breads and gave the volume of three samples of breads produced using three different methods. The context is unfamiliar and the samples were large and of different sizes. The difficulty to select a production process for maximum volume was increased by the fact that the three samples had almost the same mean and the same standard deviation. This could direct the student’s justification to other characteristics of the distribution such as spread-outness, the global form or a group of position measures such as median and quartiles. The fifth question, “Speed” (Cobb, 1999), asked if the introduction of a speed trap had changed the driving habits of 60 drivers, making them slow down. The next task, “cholesterol” (ibid.) was related to a diet program for reducing cholesterol, and presented the before and after results for 60 persons. The last situation, “T–cells”, asked students to compare two treatments for AIDS patients given the T–cell counts of a set of 46 patients receiving an experimental treatment and another set of 186 patients getting the standard treatment. Here the samples are of different sizes, which could have led to a multiplicative reasoning. The context, however, is not very familiar. Except for the fourth situation, “Breads”, we considered that the tasks were ordered by increasing difficulty.

Since we wanted to investigate the influence of the mode of presentation, the tasks were proposed in three different forms: (1) list of data, (2) graphical representation and (3) summary statistics (mean, median, quartiles, minimum, maximum, standard variation, s□□ and spread). At the end, there were three versions of the questionnaire; in each one, each task had one of the three forms and overall, each questionnaire had a variety of modes of presentation. However, because, we wanted to check if the groups were homogeneous, we used only graphical representations for questions 5, “Speed”, and 7, “T–cells”, in the three versions of the questionnaire.

One hundred and forty-one fifth year high school students (11th grade in Quebec) from five different schools completed one of the three randomly assigned versions of the questionnaire: 48 for version I and II, 45 for version III. The average age of these students was 16 years and they were enrolled in a strong mathematics course leading to college entrance.

STATISTICAL THINKING: LEVELS OF JUSTIFICATION

Based on the different descriptions of the classroom experiment of the situations (McLain, 1999; McGatha, Cobb & McClain, 1999; Cobb, 1999), in order to measure the level of statistical thinking, we established four levels of justification going from a deficient explanation to an adequate statistical perspective. In Level 0, we placed faulty or incomplete arguments such as: “We don’t know how many there are in each group so we can’t compare” or “there are more breads in the first sample”. In Level 1, we grouped the justifications using only one data point or an arithmetical partition. For example: “the battery that lasts longer is a type A battery” or “Player A has five games over 20 points”. In Level 2, we considered explanations using central tendency measures or dispersion measures that encompass all the data and, as such, characterize the distribution even if using only one measure is not generally considered sufficient for such a matter. Level 3 was reserved for justifications using a combination of central and dispersion measures, or explanations showing some multiplicative or proportional reasoning even if expressed qualitatively. It could be expressed by a phrase such as “the majority of….” or a percentage. Finally, each answer was coded “with context” or “without context”, to show if the argument drew on the context or not.

RESULTS

HOMOGENEITY

First, it was necessary to compare results of the different schools in order to see if they could be considered a homogenous group. For this, we tested the results of the level of justification and of the use of context for questions 5 and 7, which were set in only one mode in all questionnaires. These variables proved to be independent of the student's school; therefore the 141 subjects were taken as a homogeneous sample.

CONTEXT

The results show that the answers linked with the context are preponderant (72.93 %), and this is true for every task although not to the same degree (Figure 1).

These results at first sight seem to contradict the intuitive order of difficulty of the items, but other variables have to be taken into account.

The use of the context versus the mode of presentation was also tested. Where it showed a significant dependency, "Batteries", Breads" and "Cholesterol", the presentation by Summary statistics (Mode=3) seemed to reduce the exploitation of the context. For the two tasks where this dependency did not appear, "Basketball" and "Trip decision", the arguments were widely imbedded in the context (84.40 % and 94.33 %) regardless of the presentation.

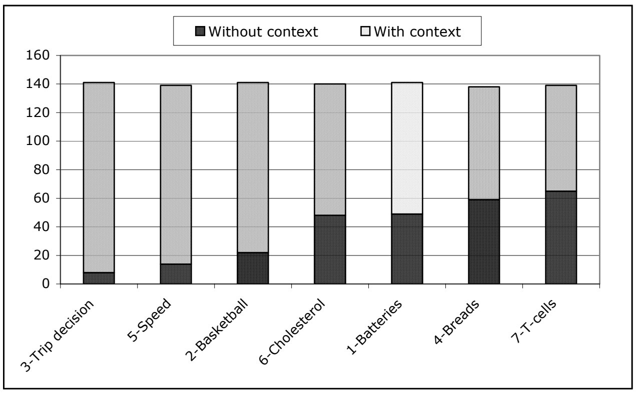

Figure 1. Use of context: Tasks are presented by increasing order of percentages for "without context", and decreasing for the other category. Columns add up to roughly the 141 participating students.

STATISTICAL THINKING: LEVELS OF JUSTIFICATION

Globally, 41.5 % of the arguments were of Level 2 in agreement with the teaching of statistics received in school that emphasize the calculation of measures presented in different grades, making it difficult for students to see the relations between them. However, if we consider that Level 3 reveals a reasoning going towards a statistical vision, almost 30 % of the answers used a combination of measures or a multiplicative reasoning (Figure 2).

In preparing the questionnaire, the hypothesis was that one of the sources of difficulty of the situations was the level of familiarity with the context. Except for question 4, the situations were ordered according to this proposition. Nevertheless, the last three problems produced more Level 3 answers. It appears that the level of familiarity with the context does not have the influence predicted. In fact, examining question by question, the results can be explained differently.

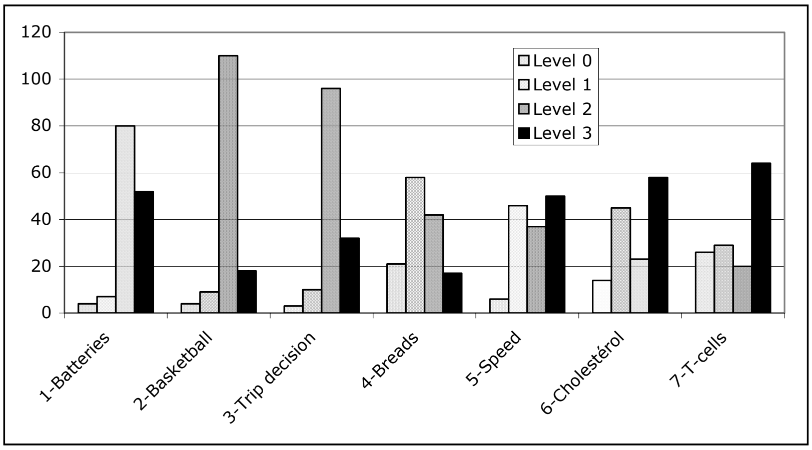

Figure 2. Number of students for each level attained for each task.

In "Batteries", the number of data was 10, making it easy to calculate of the mean that is the argument used in most students' answers. With the "Basketball" players, the number of data was eight, also easy for calculating a mean but the important difference in the dispersion of the data resulted in most students (92/141) using a "reliability" argument when talking about dispersion. The conditions were quite the same for "Trip decision" where the data for three weeks in each month were given. Here again, dispersion or, inversely, stability, was the dominant argument (93/141). In these last two situations, we can see that although the arguments referred to dispersion only, it is possible that the similarity of the central position of the data was implicitly assumed if not outwardly expressed. The "Breads" situation, an unfamiliar context, asked students to compare three large samples of different sizes (833, 939, 947) where the summary statistics were similar. Here (and in the "T-cells" problem), the samples being of different sizes explains more level 0 arguments, irrelevant answers, such as "there are more breads in this sample", "I can't compare, there are more people who follow the regular treatment…" The "Speed" and "Cholesterol" situations presented two samples of 60, and the last question, the "T-cells", had samples of different sizes (46, 186). In these three situations a remarkable proportion (≥ 36%) of arguments showed a multiplicative reasoning. It is possible that the complexity due to the larger sample size added to the less familiar context for the "Cholesterol" and "T-cells" and produced a more global perception of the distributions. The greater difficulty drove many students either to Level 3 or blocked some others at Level 0.

EFFECTS AND INTERACTION OF THE MODE OF PRESENTATION AND THE TASK ON THE SCORES

To study the effects and interactions of the mode of presentation and of the task on the level of the justification of the students, we performed a standard ANOVA analysis.

First of all, there are four levels 0, 1, 2, 3 and we used only the questions with three modes of presentation, thus eliminating tasks 5 and 7. The response variable is the scores obtained by the students for the level of justification. We thus have a randomised factorial experiment. The first factor is "Mode" of presentation, and the second is "Q", the questions. The statistical validity of the model was also established[2].

[2] Normality of the responses for each "treatment" is considered sufficient to validate the conclusions. The variances of the treatments cannot all be considered equal, but they will not substantially affect the results of the analysis, since the sample sizes of each treatment do not differ greatly. Sample sizes of the treatments are grouped around 47 (s.d.2, a maximum or 48, a minimum of 44). Anyhow, the p-values of the F-test are very small and any discrepancy from the assumptions will not result in increases of the p-values large enough to put them over the usual level of α=0,05 (Neter et al., 1996). The conclusions of the analysis reported here are statistically valid.

Table 1 *The ANOVA Table of the Analysis*

	SS	D.F.	MS	F	P
Intercept	2616.3	1	2616.23	4525.9	0.00000
Mode	7.92	2	3.96	6.8	0.00114
Q	59.61	4	14.90	25.8	0.00000
Mode*Q	21.46	8	2.68	4.6	0.00002
Error	395.97	685	0.58		

Note: The p-values are very low. The two factors and their interaction have a marked effect on the response

The factors, Mode of presentation and Question, as well as their interaction, are considered to have an influence on the Levels (very low p-values). The question now becomes which Mode and Questions are responsible for these results. Figure 3 presents graphically the "within-treatment" results, showing the means and standard error for each "Mode x Questions" (treatments), and the answer to this question becomes quite clear[3]. The main difference between the treatments is with the List of data of "Breads", whose mean score is lower than all the other questions, whatever the Mode. In this task, there were three lists of data with sample sizes of 833, 939, 947 making the interpretation difficult by simply looking at the data and the large numbers probably discouraged any calculation of measures. Within the Graphical presentation, the different effect is produced again with "Breads" that scores lower than "Batteries" (p=0.03 on the Tukey NHSD test) and marginally with "Trip decision" (p=0.08 on the Tukey NHSD test). With the Summary statistics, "Batteries" scores higher than all other questions except "Trip decision". As we can see, different questions revealed different modes as being most effective. Briefly, we can say that "Breads" was very difficult in the List mode and that in the Summary statistics mode, "Batteries" produced significantly higher level of justifications.

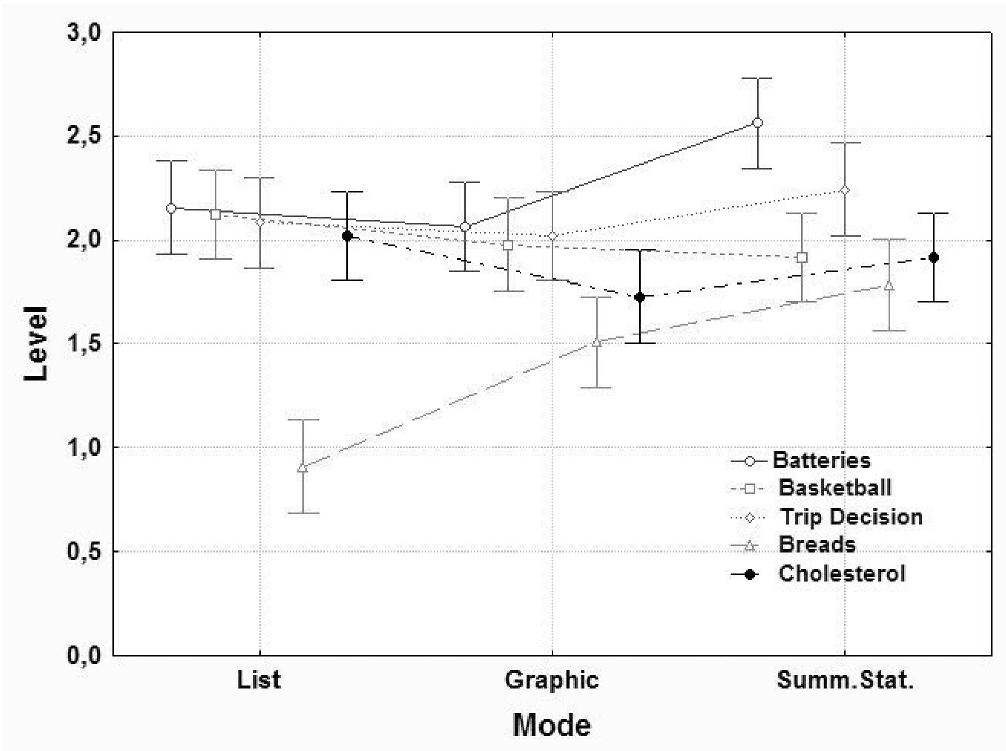

Figure 3. Mean and standard errors of the questions by modes of presentation.

[3]These intuitive results are confirmed by the usual tests: Tukey NHSD & HSD, Fisher LSD, Scheffé, etc.. They all concur.

DISCUSSION

We have found that although statistical concepts are presented in a separate and unrelated way in the Quebec high school system, some students develop a line of reasoning with a statistical perspective and are able to grasp the meaning of the data. This is progress towards understanding information based on statistics. We have shown that the mode of presentation interacts significantly with the level of difficulties of the questions, and that both of the factors have significant effects on the level of justification of the students' answers.

Furthermore, context is widely exploited and surely affects the justifications given to sustain a particular choice. For example, in the situation asking students to select a basketball player, the fact that one had a more stable performance overall was taken into account. The importance of the use of context in the students' explanations could be the consequence of the importance given to problem solving in Quebec schools where mathematical concepts (but not limited to statistics) are generally presented in a context.

These results could surely be improved. Classroom activities should include more tasks promoting discussion and asking student to compare two or more sets of data. Finally, while it is an excellent idea to start with situations where students feel confident, it is also essential to go beyond and confront them with more complex tasks to enhance the development of a higher level of reasoning, thus promoting better understanding and communication of information supported by statistics.

ACKNOWLEDGEMENTS

The authors wish to thank the referees for their very useful comments and suggestions that greatly improved the text.

REFERENCES

Cobb, P. (1999). Individual and collective mathematical learning: The case of statistical data analysis. *Mathematical Thinking and Learning*, *1*, 5–44.

Cobb, P., McClain, K., & McGatha, M. (1998). An analysis of students' statistical understandings. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Cobb, W.G., & Moore, D.S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801-823.

Gattuso, L., (2001). L'insegnamento della statistica in Québec. *Induzioni*, 2, 43-56.

Harel, G., & Confrey, J. (Eds) (1994). *The development of multiplicative reasoning in the learning of mathematics.* Albany: State University of New York Press.

Konold, C., & Higgins, L.T. (2003). Reasoning about data. In J. Kilpatrick, W.G. Martin and D. Schiffer (Eds.), *A Research Companion to Principles and Standards for School Mathematics* (pp. 192-215). Reston, VA: NCTM.

Konold, C., Pollastsek, A., & Gagnon, A. (1996). *Students analyzing data: Research of critical barriers*. Paper presented at the Roundtable in Mathematics Education (pp. 53-74). Dordrecht, The Netherlands: Kluwer.

McClain, K. (1999, April). *An analysis of the teacher's proactive role in supporting students' ability to reason statistically about data*. Paper presented at the Annual Meeting of American Education Research Association, Montreal, Canada.

McGatha, M., Cobb, P., & McClain, K. (1998, April). *An analysis of students' statistical understandings*. Paper presented at the Annual Meeting of the American Education Research Association. San Diego, CA.

McGatha, M., Cobb, P., & McClain, K. (1999, April). *Starting points of a learning trajectory*. Paper presented at the Annual Meeting of the American Education Research Association. Montreal, Canada.

Neter, J., Kutner, M.H., Natchsheim, C.J., & Wasserman, W. (1996). *Applied Linear Statistical Models* (3rd edn.). Chicago ILL: Irwin.

Raymondaud, H. (1997). Quelques pièges de la description d'une série statistique. In Enseigner les probabilités au lycée, sous la dir. de la Commission inter-IREM, Paris: Édition IREM de Reims (p. 185-212).

Russell, S.J., Schifter, D., & Bastable, V. (2002). *Developing mathematical ideas: Working with data*. Parsippany, NJ: Dale Seymour Publications.

Tukey, J. W. (1977) *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Vermette, Sylvain (2204) *Analyse et interprétation de distributions statistiques chez les élèves qui terminent leurs études secondaires.* Mémoire de maîtrise, Département de mathématiques, UQAM (Université du Québec à Montréal).