PUPILS PERCEPTION OF THE LINKS BETWEEN DATA AND THEIR ARITHMETIC AVERAGE[1]

Linda Gattuso
Université du Québec à Montréal

Claudine Mary
Université de Sherbrooke

The results presented in this paper are part of a larger study on the strategies used by pupils to solve arithmetic average problems. The arithmetic average being a widely used concept, it is important to study pupils' knowledge and their difficulties in order to improve their learning. Two questions arose from previous studies: 1)up to what point do pupils perceive the links between data and their average? and 2)does the introduction of a null datum(=0) complicate the situation? A first task asked for an estimation of the effect of the modification of a datum on the average and a second required finding the datum which would leave the average unchanged. The results tend to demonstrate that pupils have a fairly good perception of the effect of data modifications on the average with better results if the datum is zero.

Les résultats présentés sont extraits d'une étude sur les stratégies utilisées par les élèves pour résoudre des problèmes de moyenne. Vu l'utilisation répandue de ce concept, il est important de sonder les connaissances et les difficultés des élèves pour en améliorer l'apprentissage. Deux questions se sont posées à la suite d'études précédentes: Jusqu'à quel point les élèves perçoivent-ils les liens entre des données et leur moyenne? et La présence d'une donnée nulle(=0) rend-elle la situation plus difficile? Nous avons construit deux tâches dans lesquelles on demandait d'estimer l'effet du changement d'une donnée sur la moyenne et de donner la valeur qui laisserait la moyenne inchangée. Les élèves semblent avoir une assez bonne perception de l'effet de la modification d'une donnée et ceci encore mieux dans le cas d'une donnée nulle.

The arithmetic average is certainly a widely used concept and not only in the context of statistics. In fact, we can trace attempts to make many observations of the same phenomenon as far as the Babylonians (500-300 BC). It is not before the work of Tycho Brahe (16th century) that the use of the average becomes clearly distinct. However, as it is largely documented, today's students have trouble understanding this concept. In the past, many studies have stressed students' difficulties with problems involving averages, particularly when weighted averages are concerned (Pollatsek, Lima, Well, 1981; Gattuso, Mary, 1996). Clearly, this is not a simple computational algorithm, and it is not well understood.

Young children already have conceptions (or misconceptions) of representativeness, and a study from Mokros and Russell (1995) reveals that they perceive the average in five different ways: as a mode, an algorithm, a reasonable value, a midpoint or as a point of balance. However, Cai (1995), questioning sixth–grade pupils, found that although 90 % of them knew the computational algorithm for the average, less than half had reached a conceptual understanding of the concept. Other results show that 8th grade students are able to find a weighted average without previous specific instruction, and that instruction may even interfere, since older students have a poorer performance (Gattuso, Mary, 1998). This conclusion agrees with earlier findings of Pollatsek (1981), where only fourteen out of 37 college students answered correctly to a weighted average problem. Leon and Zawojewski (1990) also established that most students can understand the mean as a computational construct, but have more difficulty seeing it as a representative value. Young children seem to have a fairly good perception of the average, but this does not seem to develop into a deep understanding of the concept and difficulties show up if the problem asks for more than a simple computation, and there is no real improvement with age.

In addition, zero presents a particular problem, as in other mathematical situations, such as division by zero or 0 as an exponent. Looking into the properties of the average, a study revealed that college students with basic statistical education had a tendency to apply the four axioms which constitute an additive group to the computation of means (Mevarech,1983). Particularly, in the case where one datum is 0, some pupils consider it as the neuter element and assert it does not change the mean. Strauss and Bichler (1988) observed difficulties among 8 to 14 year-olds while looking at the development of children's concepts of the arithmetic average. Not only do children have problems understanding that the sum of the deviations to the mean is zero, but they also fail to understand that the average is representative of the values averaged and in to take into account a zero value in the computation of the average. Again, this fact seems to be independent of the age group.

RESEARCH ISSUES

In this paper we focus firstly on the understanding of the links between the data and the mean. More precisely, the issue of "representativeness" was translated in a simpler question: "Do pupils perceive that changing even one of the data affects the mean?" The second aim is to examine, in parallel, the difficulties encountered by the introduction of a datum equal to zero: "Are similar situations dealing with a datum equal to zero more difficult?"

To improve the understanding of such a concept, it is important to know how children cope with it in different situations, how much they know about it and what are the difficulties. The final aim is to investigate if pupils perceive the connections existing between the data and the average and to analyse their strategies and reasoning for eventual use in the construction of teaching interventions. Ultimately, we hope that the results of this study will help setting up future teaching experiments.

METHOD

Context of the study

Although simple arithmetic average problems are part of the elementary mathematics curriculum, it is not before the 9th grade that Québec students encounter the concept of weighed average. The results presented in this paper are part of a larger study on the strategies used to solve problems of weighted averages by children through their high-school years. A total of 638 high school students from grades 8, 9 and 10 (ages 13 to 15) participated in the study. Each student answered 5, 6 or 7 questions depending on whether they were in 8th, 9th or 10th grade respectively. A total of 24 different tasks were designed, each of which was administered to part of each year group. This paper discusses the results of 6 of these tasks.

Description of the tasks

The tasks were meant to investigate the pupils' understanding of the link between the data and the average and to see if the same situation using a datum equal to zero would pose greater difficulties. The tasks were framed into three different situations consisting of modifying the data set by: 1) adding one datum, 2) replacing one datum by another and 3) removing one datum. In each situation, two options were considered, using a non null datum or a datum equal to zero (see appendix for examples). Two questions were asked: a) Does the average increase, remain unchanged or decrease? b) If you think the average is modified, what value would leave it unchanged? Even though the first question calls for a qualitative answer and the second one does not really require computations, the numbers were chosen so as to make the calculations easy if ever the pupil found the need to do some.

We posed two hypotheses. The first one was that the question that proposed the replacement of one datum by another would be the most difficult because it required more steps that the one presenting the addition or removal of a datum. Between the latter two, adding a data should be easier. Secondly, because previous studies had shown that students do not seem to take into account zeros, it was assumed that problems involving a zero would produce poorer results.

RESULTS

Performance

Looking at Table 1, we see that for cases involving a datum different from 0, the results seem to confirm the fact that it is more difficult for the students to estimate the modification of the average if there is replacement of data; in this case, one datum smaller than the previous average was replaced by one datum greater than the original average. For problems involving addition or removal of one datum, results are higher, showing very little difference between the two.

Cases involving a datum equal to zero produce a different outcome. Our hypothesis is not supported: the case involving the replacement of one datum is no longer the most difficult one. The case where one datum is removed has the lowest results. Furthermore it seems that problems involving a datum equal to 0 are easier than previous cases involving a data ?0. Here, a datum smaller than the original average was replaced was by a datum equal to zero (also smaller than the average).

Table 1: Question a) correct answers

Data	Adding one datum	Replacing one datum	Removing one datum	Total
? 0	68/98 **69,40%**	52/107 **49,50%**	74/109 **67,89%**	194/314 **61,78%**
= 0	91/101 **90,10%**	75/104 **72,12%**	68/105 **64,76%**	234/310 **75,48%**

Looking at the results of the second question asking what should the modified datum be so the average would not change, the overall pattern is similar although the results are clearly lower. Again, the situations involving a zero seem to be easier except in the case of the removal of one datum. Let us emphasise the fact that no calculations were absolutely necessary for answering these questions. We also examined the results as a function of age group. Although there is no obvious pattern, we can say that the 8th grade students perform often better or at least as well as their older counterparts.

Table 2: Question b) correct answers

Data	Adding one datum	Replacing one datum	Removing one datum	Total
? 0	61/97 **62,89%**	46/106 **43,39%**	53/109 **48,62%**	160/312 **51,28%**
= 0	77/101 **76,24%**	61/103 **59,22%**	39/105 **37,14%**	177/309 **57,28%**

Strategies for question a)

We analysed the strategies (or reasoning) used by the students to try to understand these results. In general, in cases where the modified datum is ?0 (Table 3), we can say that the strategy used most successfully and most often by the students is to compare the modified data to the original mean. The student asserts: "a value greater than the mean is taken out, so it will decrease...". However another reasoning also focusing on the data leads to wrong answers when a datum is replaced because it says, for example: "We take out something and put back in compensates…" without any regards to the values of the data and with no reference to the mean value. They add the effects of the modified data instead of subtracting them.

Problems in which there is a zero (Table 4) are easier because the reasoning, in the case of adding or removing a datum, involves only the number of data, the total remaining the same. In the case of the replacement of one datum by another, what makes it easier is the fact that students mostly reason on the difference in the total, which in this case is the modified value: a datum ?0 is replaced by a zero, so the total decreases while the number of data remains unchanged. Here also, only one factor varies.

Table 3: Strategies: Data ?0 Question a)

Data ?0 a)	Adding one datum **?**0	Replacing one datum by ?0	Removing one datum ?0
Correct strategy mostly used **Don**	"A datum smaller than the mean is added" 45,92%	"A datum smaller than the mean is removed and a datum greater than the mean is added" 33,64%	"A datum greater than the mean is removed" 47,71%
Incorrect strategy mostly used	(Average + datum)/2 **Wwe** (without weight) 16,33%	False Compensation of effects **Fcom** 17,76% **Wwe** (as much as in Adding) 16,82%	Nothing particular

Another noticeable result is the fact that many errors are due to reasoning based on a misconception or, more accurately, a conception inadequately used. More than half of these answers are of the type: "0 is removed, it doesn't change anything…", treating 0 as the neuter element. Another erroneous reasoning is based on the conception that each datum is equal to the average. Even though it is mentioned in the problem that the datum removed is equal to zero, some pupils still think that each datum being equal to the average "Nothing changes, they all have the mean...".

Table 4: Strategies: Data =0 Question a)

Data =0 a)	Adding one datum = 0	Replacing one datum by 0	Removing one datum=0
Correct strategy mostly used	"There are more persons for the same total" **Eff** 53,4%	"The total decreases" **Tot** 38,46%	"There are less persons for the same total" **Eff** 47,0%
Incorrect strategy mostly used	Nothing particular	No explanation or answer **Nex** 12,50%	Misconception 24,8%

Strategies for question b)

The strategies change with the questions when students are asked to give the datum that would leave the mean unchanged (Table 5). In **T/n -**if you add a datum it has to be equal to the mean so the ratio (Total/nb of data does not change)– and **Mean, –**take out a value equal to the mean– we may say that the fact that "if all data should be equal, then each datum should be equal to the mean" is correctly used although it is not explicitly said. It is however difficult to distinguish between a correct answer and one misusing a conception saying that all data are equal to the mean. A great number of "no answer" or "no explanation" is also noticeable in the case of the removed and replaced data ? 0 that did not appear so much in the other problems[2].

Table 5: Strategies: Data ?0 Question b)

Data? 0 b)	Adding one datum**?**0	Replacing one datum by ?0	Removing one datum?0
Correct strategy mostly used	"If there is a person more, we add the mean to the total then it's unchanged **T/n** 24,74%	"The same as what is removed" **Com** 38,68%	"It should be the mean so it does not change" **Mean** 30,28%
Incorrect strategy mostly used	Various calculations without weight **Wwe** 19,59%	Various calculations without weight **Wwe** 7,5% And **Nex** 27,36%	No explanation or answer **Nex** 35,78%

The case of the datum equal to zero gives rise to the usual misconceptions: "Adding (or removing) 0 does not change anything..." and when the datum is replaced: "it should be the average value so everyone has the same". Many students think that every datum is equal to the mean unless otherwise specified. It is in fact possible, and also it is often said "that the mean is the value each datum would have if every data should be equal" the second part of the sentence (**if**…) seams to have disappeared.

Table 6: Strategies: Data =0 Question b)

Data = 0 b)	Adding one datum=0	Replacing one datum by 0	Removing one datum=0
Correct strategy mostly used	"It should be the mean so it does not change" **Mean** 36,63% and **T/n** 21,78%	"The same as what is removed" **Com** 50,96%	"It should be the mean because everyone has it" "Misconception" 15,24%
Incorrect strategy mostly used	Misconception 14,5%	Misconception 11,54%	Misconception 9,52%

DISCUSSION

The results call for a different point of view on the situations involved. Our hypotheses were based on previous research and on the complexity of the mathematical operation involved, but children naturally use comparison confronting the modified data with the mean or looking at what is changed, the number of data or the total. These comparisons seem easier if the datum is zero because only one factor varies.

If we look at each modification separately, we see that adding a datum, affects the mean in the same direction. If the added datum is greater than the average, the average increases and so on. The opposite is true when removing a datum. The situation were a data is replaced is not as difficult as was thought a priori, since if you look at the modification of the total while the number of data remains the same, the changes also vary in the same direction. These facts should not be presented directly but children should experience these different situations and be encouraged to estimate the results. A datum equal to zero should be treated as another datum (often smaller that the average in simpler contexts).

However, many errors are due to an incorrect application of some conception of the mean, particularly, one giving equal value to each datum. Various analogies are used in teaching but they present some limits. One example, often used in teaching the mean, is "equalising a pile" or distributing a quantity equally. The fact that all the data are not equal at the beginning should be emphasised, and probably working inversely, asking to construct a list of data given a certain mean (with or without additional constraints, like no data equal to the mean) could help the pupil to see that the data are not necessarily all equal. Every time an analogy is used, we should insist on the limits of it's validity.

Certainly, these results are incomplete and are not meant to be generalised but a majority of students do perceive the link between data and their mean, and their 'natural' strategies should be looked into and analysed for further applications in constructing teaching situations. As a matter of fact, these types of situation are scarcely used in the classroom, students should be encouraged to rely more on estimation before indulging in automatic computation.

[1] Unless otherwise specified, in this text the words "average" and "mean" are used synonymously and refer to the "arithmetic mean" or "arithmetic average".

[2] This problem was not the last one in the questionnaire.

REFERENCES

BATANERO, C., GODINO, D., VALLECILLOS,A., GREEN.D., HOLMES, P. (1994) Errors and dificulties in understanding elementary statistical concepts, *International journal of mathematics education in science and technology 25(4)*, 527-547

CAI, J..(1995) "Beyond the Computational Algorithm: Students' understanding of the arithmetic average concept". In Luciano Meira, David Carraher (Eds.), *Proceedings of the 19th International Conference for the Psychology of Mathematics Education*. Pernambuco, Brazil. (3) 144-151.

GARFIELD, J. AHLGREN, A. (1988). "Difficulties in learning basic concepts in probability and statistics: implications for research". Journal for Research in Mathematics, 19. 44-63.

GATTUSO, L., MARY, C. (1998) "Development of the Concept of Weighted Average among High-School Children". In *Proceedings of the Fifth International Conference on Teaching Statistics*. Singapour : ISI. 685-692.

LAVOIE, P., GATTUSO, L. (1998) "An Historical Exploration of the Concept of Average". In *Proceedings of the Fifth International Conference on Teaching Statistics*. Singapour : International Statistical Institute. 1051-1058.

GATTUSO, L., MARY, C. (1996). "Development of concepts of the arithmetic average from high school to university". In Luis Puig & Angel Gutiérrez (eds.),*Proceedings of the 20th International Conference for the Psychology of Mathematics Education.*. Valence, Espagne. 401-408.

LEON, M., ZAWOJEWSKI, J. (1990). "Use of Arithmetic Mean: An Investigation of Four Properties Issues and Preliminary Results". In David Vere-Jones (Ed.), Proceedings of ICOTS 3. Dunedin, New Zealand. 302-306.

MEVARECH, Z.R. (1983) "A deep structure model of students' statistical misconceptions". Educational Studies in Mathematics, 14, 415-429.

MOKROS, J., RUSSELL, S. (1995). "Children's concepts of average and representativeness". JRME, 26. (1) 20 39.

POLLATSEK, A., LIMA, S., WELL, A.D. (1981) "Concept or computation: Students' understanding of the mean". Educational Studies in Mathematics, 12, 191-204.

RUSSELL, S., MOKROS, J..(1990) "What's Typical? Children's and Teachers' Ideas

About Average". In David Vere-Jones (Ed.), Proceedings of ICOTS 3. Dunedin, New Zealand. 307-313.

STRAUSS, S., BICHLER, E. (1988). "The development of children's concepts of the arithmetic average.". Journal for Research in Mathematics Education, 19. 64-80.

APPENDIX

Replacing one data = 0

A group of eight friends empty their pockets. They have an average of 11$ each. Peter decides to take back the 4$ he put in and goes to work. At the same time, Jean-Philippe joins the group but he does not have a penny. What happens with the average?

a. 1) It decreases 2) It does not change 3) It increases

b. If you think that the average has changed, how much should Jean-Philippe have so that the average remains the same?

Removing one data ? 0

The average age of the first seven first persons attending Geneviève's party is 21 year old. When Jean-Philippe who, is 27 year old, leaves. What happens to the average age?

a. 1) It decreases 2) It does not change 3) It increases

b. If you think that the average has changed, what age should Jean-Philippe have so that the average remains the same?