Journal of the Korea Society of Mathematical Education Series D:
Research in Mathematical Education
Vol. 12, No. 1, March 2008, 49–66

韓國數學敎育學會誌 시리즈 D:
<數學敎育硏究>
제 12 권 제 1 호 2008 년 3 월, 49–66

Understanding the Arithmetic Mean: A Study with Secondary and University Students

García Cruz, Juan Antonio
Departamento de Análisis Matemático, Didáctica de las Matemáticas, Universidad de La Laguna,
La Laguna, Islas Canarias, España, Spain;
Email: jagcruz@ull.es

Garrett, Alexandre Joaquim
Departamento de Análisis Matemático, Didáctica de las Matemáticas, Universidad de La Laguna,
La Laguna, Islas Canarias, España, Spain;
Email: alexgarras@yahoo.com.br

(Received January 29, 2008; Accepted February 4, 2008)

In this paper we present a cognitive developmental analysis of the arithmetic mean concept. This analysis leads us to a hierarchical classification at different levels of understanding of the responses of 227 students to a questionnaire which combines open-ended and multiple-choice questions. The SOLO theoretical framework is used for this analysis and we find five levels of students' responses. These responses confirm different types of difficulties encountered by students regarding their conceptualization of the arithmetic mean. Also we have observed that there are no significant differences between secondary school and university students' responses.

Keywords: arithmetic mean concept, understanding, SOLO taxonomy, secondary and university education
ZDM Classification: K14, K15, C14, C15
MSC2000 Classification: 97C30, 97D70

1. INTRODUCTION AND RESEARCH QUESTIONS

Over the past few years the arithmetic mean has been widely introduced into educational curricula in various countries, given its importance in various social spheres and for the fact that it is a basic concept for the study of other subjects. Problems associated with the arithmetic mean have been worked on by students for more than 100 years (Watson & Moritz, 2000). However different research studies have shown that students' comprehension involves various types of difficulties, showing that it is not so easy for students to understand basic notions associated with this concept. These difficulties are related to different aspects: understanding of the algorithm, understanding of the concept and its properties, use of representations and language, ability to put forward arguments, among others. For example, Pollatsek, Lima & Well (1981) found that even university students can fail to take into account frequencies when solving weighted average problems.

Resulted obtained by Mevarech (1983) indicate that one possible explanation for errors detected in students' calculations of the average is that they behave as though a set of numbers together with an arithmetic mean operation were an algebraic group that satisfies the four axioms of closure, associativity, neutral element and inverse element. Results taken from the study undertaken by Strauss & Bichler (1988) with students split into four age groups of 8–, 10–, 12 and 14-year-olds who had not received any instruction in the arithmetic mean, show that students have greater difficulties understanding that the sum of the deviations of the data with respect to the average is zero, that a value zero have to be taken into account when calculating the average, and that the average is a representative of the data from which it has been calculated. Theses difficulties were also confirmed by Leon & Zawojewski (1991).

Mokros & Russell (1995) identify and analyze five basic constructions of representativity: the average as mode, the average as algorithm, the average as a reasonable entity, the average as mean point and the average as a mathematical point of equilibrium. Results from the study made by Cai (1995) with sixth-grade students showed that 90% of the students knew the mechanism "add up all and divide", which constitutes the calculus algorithm. However, only some of these students showed any proof of understanding the concept. Work carried out by Batanero, Godino & Navas (1997) identifies conceptual errors and difficulties in practical application in the knowledge of averages among 273 primary school teacher trainees. Difficulties were related to the treatment of zeros and atypical values in the calculation of averages, relative positions of averages, median and mode in asymmetric distributions, choice of the most suitable central tendency measure in a given situation, and the use of averages in the comparison of distributions. Cobo & Batanero (2004) continued their research into the difficulties secondary school students have with central tendency measures. Their results reveal that about half of the 16-year-old students they studies are capable of solving weighted average problems when putting forward suitable arguments, while only about 10% of the 13-year-old students can do this correctly. They also found errors in calculation and incorrect application of other properties.

García Cruz & Garrett (2006a; 2006b) reveal difficulties concerning students' abilities to form arguments. Students' actions were analyzed when they given open-ended questions related to multiple-choice questions. It was seen that students were inconsistent in their statements seeing that they used completely different criteria for the same situation.

Watson & Moritz (1999) analyze how students apply the average in the context of problem solving whereby sets of data presented in chart form are compared. This study was carried out with 88 students from third to ninth grade. When two charts of different sizes were compared only 2 of the 37 students from fifth to seventh grade, and 8 of the 28 students from ninth grade used the average to compare the two sets. In more complex contexts, in sets of unequal size, the responses elicited were classified into three levels: unistructural, when the responses merely referred to visual comparison; multistructural, when the students went through several steps in comparing visually or in numerical calculations that involved the average in order to compare groups, and relational, where students integrated all the proper information regarding both visual comparison as well as calculation of averages, when validating a decision as to the group they considered better. Watson & Mortiz (2000) researched longitudinally the development of the compression of the notion of averages for students from 8 to 15 years-old, from third to ninth grade. A hierarchical classification was established for students' responses at six levels of understanding. The first four levels describe the development of the concept of average taken from colloquial ideas in procedural and conceptual descriptors which induced a central position measure from a set of data. The two highest levels were represented by the transfer of this understanding to one or more applications in tasks involving the solution of problems regarding the inversion of the calculus algorithm and calculation of a weighted average. The results showed that for many young students the average is merely a value in the centre of distribution (an idea close to that of the median). Hardly ever do these children relate the word "average" to the mode and even less to the arithmetic mean. The researchers show that students' responses to a certain degree tend to ascend in category with regard to the level of understanding as the students proceed through the education system.

In this paper we attempt to contribute with some of ours findings from a study carried out with Angolans' students. Providing that this information can be of value in terms of intercultural experiences, an aspect claimed in educational research (Shaughnessy, 1992).

This work is part of a broader study that we have been carrying out on the arithmetic mean. In this work we give the results obtained when going deeper into analysis of different levels of cognitive development that can be seen when students are learning the arithmetic mean, and set up a hierarchical structure of the results taken from analysis of students' responses to two of the questions which assesses: knowledge of the arithmetic mean as the best estimator of a measure when there are errors in measurement; the influence of atypical values in calculation of the arithmetic mean.

We aim to answer some research questions related to the understanding of the concept, such as:

- What are the predominant levels of understanding demonstrated by students in ccordance with the responses elicited from the tasks set?
- What differences in levels of understanding can be found between secondary school and university students?

We have made a hierarchical characterization because it allows differentiating the types of responses given by students, and also helps to reflect whether the responses are in accordance with the level demanded in the task, according to the educational year the students are at, and consequently programmed teaching situations that nurture students' thinking. In addition, this process relates the outcomes of learning with the objectives initially set, supplying us with information that help evaluate feedback from the teaching and learning process (Biggs & Collis, 1982). From our bibliographical review, only in the work made by Watson & Moritz (1999; 2000) do we specifically find this type of analysis, although their research is exclusively concerned with primary and secondary school students, while the tasks they analyze are limited to situations wherein two sets of data given in chart form are compared and are restricted to the use of the terminology of average in everyday situations: one being the number of hours dedicated to watching television, and the other to the number of children in a family. In our study we look at other aspects of the arithmetic mean in different contexts, while our study is also undertaken with university students, something which has not been undertaken before in this type of study.

CONCEPTUAL FRAMEWORK

According to Schroeder (1987), mathematics educators have been concerned with students' understanding for many years, but in the last few decades activity and progress in this field have been substantial. This idea is corroborated by Hiebert & Carpenter (1992) who state that there is a general consensus within the mathematics educational community that students should understand mathematics. For Hiebert & Carpenter, the aim of much research and implementation of work in this field of teaching have been related to the promotion of meaningful forms of learning. They also point out that the problem of understanding goes beyond the boundaries of mathematic education.

Furthermore, examining the works of various researchers, such as Skemp (1976, 1979), Hiebert & Lefevre (1986); Sierpinska (1990); Sfard (1991) & Godino (1996) among others, we can find numerous models and conceptual frameworks built around understanding. This demonstrates that researchers approach the idea of "understanding" from different points of view, although different analyses share common elements.

In our work we attempt to build on the bases of the neo-Piagetian theory described by Biggs & Collis (1982; 1991). This theory constitutes a hierarchical model for the study of students' development throughout the learning process, based on a set of tasks limited to a given domain. Various levels can be differentiated within this conceptual framework: The prestructural level is associated with initial responses and shows that learning is of a very low level with respect to the level of abstraction the task demands. The student is vague in his or her responses or is attracted by irrelevant aspects of the task. At the unistructural level can be found those students who are able to focus on the proper domain of the task but who take into account only one aspect of that domain. The multistructural level is characterized by the fact that the student learns more and more suitable or correct aspects of the task but is unable to integrate them. At the relational level the student integrates the various parts and with them completes a coherent and meaningful structure.

Taking into account all of the above, we decided to work with the original model proposed by Biggs & Collis (1982; 1991) establishing five levels of understanding: prestructural, unistructural, multistructural, transitional between the multistructural and the relational, and finally the relational. This classification is based on the analysis carried out of the responses given by the students taking part in our experimental study, a description of which will be given in the section on analysis and classification of responses in this work.

METHODOLOGY

Sample

The study was undertaken with 227 students. Of these, 130 were secondary school students in the last but one year before university entry and who were aged between 16 and 21 years old. The remaining 97 subjects were studying at a university Education Faculty, 31 of who were majoring in Mathematics while 66 were in Education. These students were aged between 22 and 49 years old. The secondary school students and those university students specializing in Teaching had received instruction in Descriptive Statistics in the previous year academic year, while those students majoring in Mathematics had studied Inferential Statistics.

Data collection instruments

The data for this study were collected through the administration of a questionnaire made up of nine problems, some of which were formed by subsections, making a total of

13 items. The questionnaire is part of a wider study we are carrying out on the understanding of the concept of arithmetic mean. Here we analyze the responses given to two of these problems: One involve an open-ended question and the other a multiple-choice question.

Problem "Time taken over 100 metres": *When asked by their PE teacher, 10 students independently and simultaneously recorded the time taken by another student to run 100 m. The times recorded (in seconds) were the following:*

15.05; 14.95; 15.05; 15; 10; 15; 14.90; 15; 14.95; 15

What time should the teacher consider as the estimation of the real time taken by the student, and why?

This is an open-ended problem we took from a multiple-choice question designed by Garfield (2003), our multiple-choice question being called "In a science class". Our aim here was to compare how students act when faced with this open-ended problem ("Time taken over 100 metres") and how they then act when faced with a multiple-choice question ("In a science class") in order to examine the results obtained by García Cruz & Garrett (2006a; 2006b) in relation to the inconsistent way students act when faced with two different but interrelated types of questions. The task assesses use of the average as the best estimator of an unknown value when there are measurement errors, the influence of atypical values in the calculation of the average, as well as students' possible confusion between the average and other central tendency measures.

Problem "In a science class": *Nine students weighed a small object with the same instrument. The weight recorded by each student (in grams) is as follows:*

6'2	*6'0*	*6'0*	*15'3*	*6'1*	*6'3*	*6'2*	*6'15*	*6'2*

The students want to find out as accurately as possible the real weight of the object. Which of the following methods would you recommend?

(Mark only one of the following answers)

a. Use the number repeated most, which is 6.2.
b. Use 6.15 as this is the recording with most decimal places.
c. Add up the 9 numbers and divide by 9.
d. Discard the number 15.3 and add up the other 8 numbers and divide by 8.

ANALYSIS AND CLASSIFICATION OF THE RESPONSES

Students' answers were arranged into categories. The classification follows the theory set out in Biggs & Collis (1982; 1991). The levels of understanding encountered in each problem can be seen more clearly in the examples we provide below which illustrate the types of responses given by students.

Problem "Time taken over 100 m"

- *Prestructural responses (P)*

Responses classified at this level were as follows:

- 10 seconds should be considered as this is the time of the runner who came first; The real time run by the student should be 10 seconds because that would be the ideal time for 100 metres;
- 10 seconds should be considered as an estimate of the real time because 10 x 10 = 100;
- To estimate the time the teacher should consider 14.95 seconds. After running 100 metres we must take into account the student's intentions, effort, courage, among other factors;
- The time to be considered is 10 seconds;
- 10 seconds should be considered. As the teacher tries to evaluate the time, this cannot be more than 10 seconds;
- It's 15 seconds, because it depends on a student's fitness and ability to do this type of sport or activity over a distance of 100 metres, and the student has to be well trained;
- The time the teacher should consider as an estimate of the real time is 15 seconds.

- *Unistructural responses (U)*

Responses classified at this level were as follows:

- 10 seconds should be considered as it's the shortest time;
- The teacher should take 10 seconds as 10 is the number closest to the current world record;
- The shortest time should be considered the best estimate;
- The longest time should be considered as the seconds pass quickly;
- The teacher should consider 10 seconds because that's the shortest time and we can say that the student is faster;
- 10 seconds because in a 100 metre race the minimum time is 9.60 seconds, and for me that would be almost impossible and if the teacher wants to consider a time, it has to be 10 seconds. Also 10 seconds is the closest time;
- 15.05 seconds should be considered because it's repeated twice;
- The time the teacher should consider as an estimate of the real time is 15 seconds

because no student is slower than 15.90.

- *Multistructural responses (M)*

Responses classified at this level were as follows:

- The time the teacher should consider as an estimate of the real time run by the student is 15 seconds because this time is the most repeated in the data;
- The time the teacher should consider as an estimate of the real time is 14.90 seconds because this is the times closest to the average of 14.49;
- The time the teacher should consider is 15 seconds because of the 10 students 4 record this time;
- None of the times because the data are given as a whole. The times need to be grouped together and then find the average time run by the student can be found;
- The median should be used, as this is the value between the fifth and sixth positions.

Also classified at this level were those justifications where the calculus algorithm for the average was used but where the complete data or total sum were not correctly identified.

- *Transitional responses (T)*

Of those responses classified at this level were responses that included the average to estimate the unknown value in the situation given. However, the students giving these responses fail to realize that there is anomalous data in the set of values that must be discarded before carrying out the calculation; consequently, the students add up all the values and divide by the total number of observations, as illustrated in the responses shown below:

- The sum of the times recorded is 144.9. Dividing this we get 14.49. So the time is 14.9 seconds.
- The time the teacher should take as the estimate of the real time taken by the student is 14.49 because

$$\frac{15'05+14'95+15'05+15+10+15+14'90+15+14'95+15}{10}=\frac{144'9}{10}=14'49\,.$$

In some responses students mistakenly round up the average they have calculated for all the values to a whole number, as shown in this example:

- The value 14.49 ≈ 15 should be considered because, to find the average, you need to add up and divide by the amount of numbers added together.

Here are other examples of responses classified at this level:

- The teacher should consider the time of 14.90 as the estimation because this is the number closest to the average of 14.49;
- The time the teacher should consider is 15 seconds because

$$\bar{x} = \frac{15'05 + 14'95 + 15'05 + 15 + 10 + 15 + 14'90 + 15 + 14'95 + 15}{10} = 14'49 .$$

The average value can't be found at these intervals, so it would be 15 seconds as this is the real time recorded;

- Formula: $\bar{x} = \frac{15'05 + 14'95 + 15'05 + 15 \times 4 + 10 + 14'90 + 14'95}{10}$; $\bar{x}$ =75'6 s.

This is the time the teacher should take as an estimate.

• *Relational responses (R)*

At this level we classify those responses that show that, in order to estimate the value sought, the arithmetic mean must be calculated from the data given, while discarding the value of 10 seconds, this being an anomalous observation within the set of data. However, we could not find any response at this level with regard to the problem under discussion, as will be shown in the results.

We can now look at another example in a multiple-choice problem which is closely related to the above problem with respect to the objectives.

Problem: "In a science class"

As this is a multiple-choice question classification was centred on the distracter evident in each option. Initially only responses at the multistructural, transitional and relational levels were examined. Responses at other levels could be included in a future version of the questionnaire in order to complete the range of choice.

Classification was as follows:

• *Multistructural responses (M)*

Those responses in which students chose the option that says "use the most repeated number, which is 6.2". Our interpretation is that this infers the mode as the response, this measure not being best estimator for the situation given, and the option "use 6.15, as this is the value with most decimal points", inferring that the student confuses greater accuracy with more decimal points.

• *Transitional responses (T)*

Those responses where the subject chooses the option "add up all the values and divide the total by the total number of data". Here we interpret that the student understands the average to be the best estimator of an unknown value when there are measurement errors, but fails to pay attention to the observation that is different from the others. Estrada (2002) notes that this response might be considered partially correct, as the student uses the idea of the average as the best estimator, but fails to take into account the lack of solidity of this measurement in the face of atypical values.

- *Relational type response (R)*

Those responses where students selected the option that implied discarding the value 15.3, adding the other 8 numbers and dividing by 8. In other words, the student uses the average as the best estimator of an unknown value when there are measurement errors and also appears to realize that in these situations the extreme values need to be treated.

Finally, students who left questions blank were classified as NA (No Answer).

RESULTS AND DISCUSSION

Results are set out in accordance with each problem. We give the absolute frequencies and percentages according to the level of understanding and according to the educational sector the students belong to (secondary school or university). From the data we collected we draw our conclusions based on students' actions.

In Table 1 we set out the data related to the problem "Time taken over 100 metres". The task is taken from the set of problems and situations that emerges from the concept of arithmetic mean (Batanero & Godino, 2001). In addition to discovering that the arithmetic mean is the best way of estimating an unknown quantity when there are measurement errors, analysis should also be made to find out whether any anomalous value was not introduced into the process of data collection.

Table 1: Results for the problem "Times taken over 100 m"

	Secondary		**University**	
	N	**%**	**N**	**%**
NA	16	12.3	10	10.3
P	14	10.8	11	11.3
U	15	11.5	6	6.2
M	47	36.2	36	37.1
T	38	29.2	34	35.1
R	0	0	0	0
Total	130	100	97	100

The data show there was no response for this problem from either secondary school or university students which could be classified at the relational level. Recognizing an atypical observation and then discarding it in order to calculate the average, when using this measure as the best estimator of the time taken by a student over 100 metres, was completely unknown. We believe that, as a consequence of this, most responses were found firstly at the multistructural level (47 (36.2%) secondary school students and 36 (27.1%) university students), and secondly at the transitional level where we find 38 (29.2%) answers from secondary school students and 34 (35.1%) responses from university students. First of all, at the multistructural level, the responses show that the students have some knowledge as a result of the instruction they have received but that they fail to associate this knowledge with the objective of the task. Some students write the calculus algorithm for the average incorrectly; others give the most frequent value, in other words, the mode, while others use the median as the best estimator, even if this is badly defined. Secondly, at the transitional level, the responses show a degree of knowledge close to the relational level, although they lack certain refinements. The responses show that the students recognize that the average is the best estimator in the given circumstances, but that they also fail to achieve a more refined appreciation of the data supplied in the problem which should translate into the discarding of the atypical value before calculating the average in order that their reasoning could be placed at the relational level.

Unistructural level responses were more prevalent among secondary school students, 11.5%, as opposed to 6.2% among university students. Some of the answers at this level derived from students' own experiences, while other responses centred on only one aspect of the data or only took into account one concrete attribute, the students believing this sufficient for the situation. Finally, it appears that the students indeed wish to answer with something and find a response, but were not concerned that their response contains the argument sought.

The results discovered for the responses at the prestructural level are very similar for both secondary school and university students: approximately 11% of responses. In general, students' arguments evidenced gaps in understanding the problem.

Overall, the results show that the students did not know how to use the average as the measure which best functions as the estimator of a set of data when there are errors in measurement and the influence that extreme values have on calculations. Although one of the main difficulties students have in calculating the arithmetic mean is the recognition of atypical values, we were very surprised to find that not even one of the students taking part in the questionnaire realized that there were anomalous data in the set of figures they were given, especially when the diversity of life experiences held by the students, particularly university students, should lead them to realize how very improbable it would be to find a student capable of running 100 metres in 10 seconds when even top professional athletes cannot even do this time.

Difficulties with atypical values were also observed in the study undertaken by Batanero, Godino & Navas (1997) with primary school trainee teachers. The researchers explain that these difficulties are the result of the lack of contexts provided in the teaching of statistics encountered by students and the gaps in knowledge of what they have learnt. We agree with this reasoning: we observed through a survey carried out among teachers of Statistics that, in general, knowledge of the influence of values, especially regarding whether they should be discarded or not when calculating the average, is not explicitly covered.

In the problem "In a science class", the results were somewhat different when compared with the previous problem although the objectives to be assessed were the same. We think that this is because the problem demands a response to a multiple-choice question, while the first problem was open-ended. The data found are summarized in Table 2.

Table 2: Results of the problem "In a science class"

	Secondary		University	
	N	%	N	%
NA	3	2.3	0	0
P	0	0	0	0
U	0	0	0	0
M	47	36.2	44	45.4
T	61	46.9	41	42.3
R	19	14.6	12	12.4
Total	130	100	97	100

The options for responses to the problem only correspond to the multistructrual, transitional and relational levels. Although each alternative itself provided clues to the most feasible response, few students chose the relational level option: only 19 (14.6%) secondary school students and 12 (12.4%) university students. In our study, then, there is little difference in terms of reasoning between the two groups when it comes to solving this problem. The relational level was linked to the option that suggested that they discard the value 15.3 and calculate the average only with the other 8 numbers in order to work out as accurately as possibly the real weight if a small object is measured with the same instrument.

Responses at the transitional level show that students realize that the arithmetic mean is the best estimator when there are errors of measurement but fail to taken into account the atypical value observation. 61 (46.9%) secondary school students and 41 (42.3%) university students showed reasoning at this level.

At the multistructrual level, 47 (36.2%) secondary students and 44 (45.4%) university students chose the relevant option for this level. We associate this level with responses corresponding to the most repeated value or to the use of the value with most decimal places.

In order to examine some of the differences found in the data between students' actions when answering open-ended questions and their actions regarding multiple-choice questions, seeing that both these types of questions basically covered the same objectives, we compared the results obtained at each educational level.

In Table 3 we give the results for secondary school students.

Table 3: Comparison of results for the open-ended problem "Time taken over 100 metres" (in rows) and the multiple-choice problem "In a science class" (in columns) with secondary school students

		Problem "In a science class"				Total
		NA	**M**	**T**	**R**	
Problem "Time taken over 100 metres"	**NA**	0	8	6	2	16
	P	1	4	9	0	14
	U	1	9	4	1	15
	M	1	22	19	5	47
	T	0	4	23	11	38
Total		3	47	61	19	130

First of all, we analyze the students' actions for the open-ended problems based on their responses shown for the multiple-choice problem. In our data, we can see that of the 19 students who chose the relational level response for the multiple-choice problem (column R), when answering the open-ended question (rows), 11 put forward an argument that was classified as transitional between the multistructural and relational levels. For the remaining students, 5 wrote justifications at the multistructural level and 1 at the unistructural level.

Also, of the 61 students whose responses were categorized as transitional for the problem "In a science class", only 23 gave a response at the same level for the open-ended problem. Most students gave answers at lower levels, while 6 students did not even answer this problem.

Now, making the reverse analysis, we can see that of the 38 students who in the open-ended problem gave a response categorized as transitional between the multistructural and relational levels (rows), 11 (about 29%) of them chose the option categorized at the relational level for the multiple-choice question, 23 (about 61%) chose the option at the transitional level, and 4 (10.5%) chose the option classified at the multistructural level.

The actions undertaken by the trainee teachers were similar to those described above, as shown in Table 4.

Table 4: Comparison of results for the open-ended problem "Time taken over 100 metres" (in rows) and the multiple-choice problem "In a science class" (in columns) with university students.

		Problem "In a science class"			Total
		M	**T**	**R**	
Problem "Time taken over 100 metres"	**NA**	4	4	2	10
	P	3	6	2	11
	U	2	2	2	6
	M	25	7	4	36
	T	10	22	2	34
Total		44	41	12	97

Here, we can see that of the 12 students who gave responses classified at the relational level for the multiple-choice question ("In a science class") 2 gave answers at the transitional level for the open-ended problem, while 4 were at the multistructural level, 2 at the unistructural level and 2 even at the prestructural level. Another 2 students failed to give an answer.

Also, of the 41 students who selected the option classified at the transitional level for the multiple-choice question (columns), only a little more than half (22) showed the corresponding type of reasoning when answering the open-ended question.

Analyzing the actions of students in multiple-choice questions after giving an answer to the open-ended question, we can see that of the 34 students who gave responses at the highest level (Transitional) for the open-ended problem, 2 of them chose the relational type option for the multiple-choice question, 22 chose the option at the transitional level and 10 at the multistructural level.

These inconsistencies lead us the think that some students answered the multiple-choice question without taking into account any reference knowledge, as for similar situations they use different arguments. This reflects the results already described in García Cruz & Garrett (2006a; 2006b).

CONCLUSIONS

Our study has allowed us to observe various types of difficulties students face regarding the arithmetic mean. The results obtained show that students are not familiar with the notion of an atypical value, and consequently have no idea how to react when they encounter such data in their calculations of the arithmetic mean. As we have shown there were no responses at the relational level for the open-ended question and that, furthermore, there were very few responses at this level for the multiple-choice questions in spite of the fact that the options supplied gave a clue that might orientate the student.

We can also note that some students fail to recognize that the average is the central tendency measure that can best be used as the estimator. In situations where they use this measure, students preferred the mode which is not the best estimator (Batanero et al., 1997) or they gave a value that did not fulfil the demands.

In general, the results show that the students surveyed are not familiar with some of the main properties of the arithmetic mean, in spite of its elemental nature. What is more noteworthy: there were no significant differences between secondary school students and university students with respect to the levels of interpretation observed in our study. Obviously, we expected a greater level of performance from the university students as they are studying at a higher level and supposedly benefit from greater maturity and experience. Some of these students will even have to teach these contents themselves in the future. In this case, our results do not agree with those given by Watson & Moritz (1999) who state that students' level of understanding goes up as they progress through the educational system.

In our opinion, difficulties arise because students are not familiar with the conceptual aspects of the average. They study the calculus algorithm and use the proper representations, but do not expressly cover the most tangible properties. As we have discovered in this study, student can obtain the arithmetic mean for a given set of data, if this is what is asked of them, because most of them know the calculus algorithm. However, this procedural knowledge is not linked to conceptual aspects, evidencing the conclusion highlighted by Mokros & Russell (1995) regarding some students' poor conceptual understanding in that they conceive the average as a pure algorithm.

In consequence, we agree with the propositions described earlier in this paper regarding the importance of contextualizing the tasks in the process of teaching and learning. We also believe that it is important to diversify the contexts presented in problems set in the classroom; in other words, teachers should not insist on the usual contexts or those most often presented to the student, as this restricts students' thinking. We believe that diversification of the contexts helps students perceive and analyze different situations outside their normal environment and furnishes them with the bases to face up to different types of problems.

REFERENCES

Batanero, C. & Godino, J. D. (2001). Developing New Theoretical Tools in Statistics Education Research. In*: Proceedings of the 53rd Session of the International Statistical Institute (ISI) held at Seoul, Korea, August 22–29, 2001, Bulletin of ISI* (Tome LIX, Book 2, 137–142), ISI, Seoul. http://www.stat.auckland.ac.nz/~iase/publications/4/250.pdf

Batanero, C., Godino, J. D. & Navas, F. (1997). Concepciones de Maestros de Primaria en Formación sobre los Promedios. *VII Jornadas LOGSE: Evaluación Educativa*, 310–324.

Biggs, J. B. & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York, NY: Academic Press. MATHDI **1985c.**03244

_____ (1991). Multimodal Learning and the Quality of Intelligent Behavior. In: H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57–76). Hillsdale, NJ: Erlbaum.

Cai, J. (1995). Beyond the computational algorithm. Students' understanding of the arithmetic average concept. In: L. Meira & D. Carraher (Eds.), *Proceedings of the XIX Conference on the Psychology of Mathematics Education* v.3 (pp. 144–151). Universidad de Pernambuco.

Cobo, B. & Batanero, C. (2004). Numerical arguments on mean problems (Razonamiento numérico en problemas de promedios). *Suma* **45**, 79–86. MATHDI **2004d.**03616

Estrada, A. (2002). *Análisis de las actitudes y conocimientos estadísticos elementales en la formación del profesorado*. Tesis doctoral. Universidad Autónoma de Barcelona.

Gal, I. (1995). Statistical Tools and Statistical Literacy: The Case of the Average. *Teaching Statistical* **17(3)** , 97–99. MATHDI **1995f.**03878

Garcia Cruz, J. A. & Garrett, A. J. (2006b). Students' actions in open and multiple-choice questions regarding understanding of averages. In: Novotná, J., Moraová, H., Krátká, M. & Stehlíková, N. (Eds.). *Proceedings 30th Conference of the International Group for the psychology of Mathematics Education*, Vol. 3 (pp. 161–168). Prague: PME.

Garcia Cruz, J. A. & Garrett, A. J. (2006a). On average and open-end questions. *Proceedings of the 7 th International Conference on Teaching Statistics*. Salvador de Bahia, Brazil.

Garfield, J. B. (2003). Assessing Statistical Reasoning. *SERJ - Stat. Edu. Res. J.* **2 (1)**, 22–38. MATHDI **2003d.**03467

Godino, J. D. (1996). Mathematical concepts, their meaning and understanding. In: L. Puig et. al. (Eds.), *20th Conference of the International Group for the Psychology of Mathematics Education (PME 20). Proceedings. Vol. 2* (pp. 417–424). Universidad de Valencia. MATHDI **1999b.**00783

Hiebert, J. & Carpenter, T. P. (1992). Learning and teaching with understanding. In: D. W. Grouwns (Ed.), *Handbook of research in teaching and learning of mathematics* (pp. 65–97). New york: Macmilan.

Hiebert, J. & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In: J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Erlbaum.

Leon, M. R. & Zawojewski, J. S. (1991). Use of the arithmetic mean: An investigation of four properties. Issues and preliminary results. In: D. Vere- Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 302–306). Voorburg, Holanda: International Statistical Institute.

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in mathematic* **14**, 415–429. MATHDI **1984e.**11098

Mokros, J. & Russell, S. J. (1995). Children's Concepts of Average and Representativeness. *Journal for Research in Mathematics Education* **26(1)**, 20—39. MATHDI **1995f.**03881

Pollatsek, A.; Lima, S. & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics* **12(2)**, 191–204. MATHDI **1981k.**00533

Schroeder, T. L. (1987). Students' understanding of mathematics: A review and synthesis of some recent research. In: J. Bergeron; N. Hercovics, & C. Kieren (Eds.), *Psychology of mathematics Education* XI, Vol.3 (pp. 332–338). Montreal: PME.

Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics* **22(1)**, 1–36. MATHDI **1992a.**03697

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and Directions. In: D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465–494). Macmillan, New York.

Sierpinska, A. (1990). Some remarks on understanding in mathematics. *For the Learning of Mathematics* **10(3)**, 24–41. MATHDI **1992c.**01326

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching* **77**, 20–26. EJ154208

_____ (1979). *Intelligence, Learning, and Action. A foundation for theory and practice in education*. Chichester (etc.): John Wiley &Sons.

_____ (2006). Relational understanding and instrumental understanding. Math. Teach. M. Sch. **12(2)**, 88–95. MATHDI **pre**02370811, EJ765695

Strauss, S. & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education* **19 (1)**, 64–80. MATHDI **1988d.**01776

Watson, J. M. & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Education Studies in Mathematics* **37**, 145–168. MATHDI **1999e.**02977

_____ (2000). The Longitudinal Development of Understanding of Average. In: Brian Greer (Ed.), *Statistical thinking and learning. A special issue of 'Mathematical Thinking and Learning'* vol. 2 (1/2) (pp. 11–50). MATHDI **2001f.**05345