MEANING OF THE DISPERSION AND ITS MEASURES IN SECONDARY EDUCATION

Juan Ortega Moya
UNED, Spain
Antonio Estepa Castro
University of Jaén, Spain
JOMOYA@telefonica.net

In this paper we present an onto-semiotic macroscopic analysis of the measures of dispersion: range, interquartile range, average deviation, variance, standard deviation and coefficient of variation by following the theoretical framework of the Theory of Semiotic Functions. This research has been carried out with a sample of textbooks from the most representative publishers used by Spanish second-cycle Secondary students of 15 and 16 years of age. The paper finishes by presenting several useful conclusions for the planning of the teaching process and for the research on the issue.

INTRODUCTION

During the first years of instruction, the teaching of statistics focuses on the topics of location and variability. Although one admits that in the last decades of the XXth century researchers on statistical education have paid great attention to measures of centralization (Moore, 1990; Shaughnessy, 1997), tendencies are changing and at the beginning of the XXIst century variation is seen as the be all and end all of statistics (Shaughnessy and Ciancetta, 2001). According to Moore (1990), "*the ability to deal intelligently with variation and uncertainty is the goal of instruction about data and chance.*"

Convinced of the importance of variation, we endeavours to inquire into the nature of measures of dispersion, to describe the obstacles that students come across when trying to understand these measures and to look for adequate sequences of instruction in order to improve both the capacity to teach and learn them.

In this study an epistemological analysis has been carried out of measures of dispersion. We have based this analysis on a selection of textbooks of the second cycle of secondary education (15 and 16 year-old students). By looking at different textbooks one is provided with a variety of the components of the meaning of measures of dispersion from which one can conclude that some of these present certain tendencies and deficiencies (Cobo and Batanero, 2004). They also allow us to see how to guide and design the study of training processes.

THEORETICAL FRAMEWORK

This study comes within the Ontological-Semiotical Approach of mathematical instruction and cognition (OSA), also known as Theory of Semiotical Functions (TSF), developed by J. D. Godino and his collaborators, who have been developing it for more than ten years (Godino, 2003). The OSA adopts anthropological, ecological and systemic theories about mathematics; it considers mathematics as the outcome of human activity employed to resolve problematic situations, which could be either external or internal to mathematics itself. As a result of this activity mathematical objects emerge as entities that are used to resolve problematic situations. In the OSA, the analysis of mathematical activity is carried out by introducing six types of primary entities:

- *Situations*: Problems which are more or less open and that induce one to mathematical activity. An example would be to determine the maximum error of a set of measures.
- *Actions*: What subjects carry out when they try to resolve problems and what become routine actions with the practice (operations, algorithms, procedures). It is common practice in statistics to calculate a measure of dispersion for a set of data.
- *Language*: Ostensive elements of mathematical activity, words and expressions, symbols, formulations, equations, graphics, tables, ...
- *Concepts*: Definitions, notions and mathematical ideas.
- *Properties:* Characteristics of mathematical objects.

- *Arguments:* Reasoning that relates the previous elements of meaning and justifies their properties.

When a process of instruction is planned on a mathematical subject, one must begin by specifying "*what this topic means as far as mathematical and didactic institutions are concerned*" (Font, 2004). This means that the epistemical dimension of the object in question is drawn up. In order to do this, one has to look at mathematical textbooks, curricular orientations, historical sources and, in general, what "experts" consider to be the operative and discursive practices inherent in the object. All this amounts to the *institutional meaning of reference* of the object. In the OSA does not exists a only institutional meaning of reference. On the contrary, a variety of meanings are possible depending on the institution which imposes these problems and where the activities are being carried out to solve them.

At Secondary Education, curricular approaches, textbooks and the "knowledge of teachers" describe functionality, an operating strategy and a representation of measures of dispersion which characterise their meaning at this educational level and which may prove different to those held by a university institution (Estepa and Ortega, 2006).

METHOD

The aim of present study is to characterise the institutional meaning of measures of dispersion, Range (R), Interquartile Range (IR), Average Deviation (AD), Variance (V), Standard Deviation (S) and Coefficient of Variation (CV), which are introduced in Spanish textbooks in the 3rd and the 4th years of Compulsory Secondary Education.

In order to carry out this analysis, a selection of textbooks from the most widely used publishers in Spain have been chosen (see appendix). Specific issues on statistics have been selected from each textbook and within these issues the sections which refer to measures of dispersion have been analysed, as well as the activities proposed in examples and exercises. Every section has been divided in units and sub-units, which were in turn assigned to the aforementioned categories of primary entities (situations, actions, language, concepts, properties and arguments).

RESULTS

Situations

S1. Variation by ranges: Finding the maximum difference in a set or subset of data.
S2. Variation by deviations: Measuring the variation around a centre.
S3. Global comparisons. Two kinds:
 S3a. Comparing the variation of two or more distributions measured in the same magnitude.
 S3b. Comparing the variation the two more distributions measured in different magnitude.
S4. External local comparisons: Comparing the relative position of data in different distributions
S6. Inverse problem: Generating data, identifing graphs, tables,...from information about its variation.
(See Estepa and Ortega (2006) for the codification and other kinds of situations.)

Table 1: Frequency of the situations found and the publishers in which they appear

		SITUATIONS					
		S1	S2	S3a	S3b	S4	S6
PUBLISHING HOUSE	Anaya	5	56	16	1	0	0
	Santillana	24	47	22	0	0	0
	Bruño	15	19	9	0	0	1
	McGrawHill	11	19	11	0	0	0
	Edelvives	13	56	2	6	0	0
	Oxford	11	29	10	0	0	0
	SM	9	17	5	0	0	2
	Casals	10	42	10	0	1	2
Total situations		98	285	85	7	1	5
Total Publishing Houses		8	8	8	2	1	3

We can deduce, observing Table 1, that the situations: S1, S2 and S3a are present at all publishers and the rest almost are not used.

Actions

We have analyzed the following actions (techniques, algorithms, procedures, ...):

A1. Calculating measures of spread of a series of data (without tabulation).
A2. Calculating measures of spread of tabulated data (not grouped).
A3. Calculating measures of spread of tabulated data, grouped in intervals.
A4. Calculating measures of spread with calculator and/or computer.
A5. Calculating measures of spread of transformed data (brief calculation).
A6. Calculating measures of spread of data displayed graphically.
A8. Calculating measures of dispersion without the data (from information based on properties, definitions, formulas, etc.).
A9. Making inverse calculations with dispersion measures.
A10. Calculating and interpreting the interval $\left(\bar{x}-k\sigma,\ \bar{x}+k\sigma\right)$ and the percentage of data that it contains.
A11. Representing graphs that contain information on dispersion.
A12. Standardizing data.

Table 2: Frequency of actions analyzed and the publishers in which they appear

PUBLISHING HOUSE	ACTIONS										
	A1	A2	A3	A4	A5	A6	A8	A9	A10	A11	A12
Anaya	30	16	17	6	0	4	9	0	3	0	0
Santillana	61	17	2	10	0	9	0	0	4	1	0
Bruño	29	0	7	2	0	0	0	0	0	2	0
McGrawH	11	9	14	7	0	0	0	0	0	0	0
Edelvives	81	23	9	1	0	2	0	0	0	0	0
Oxford	25	10	9	5	0	14	0	0	2	0	0
SM	17	3	7	1	0	0	0	2	4	0	0
Casals	18	16	4	18	1	6	0	2	9	0	1
TOTALS	272	94	69	50	1	35	9	4	22	3	1

Other kinds of actions are been found in Estepa and Ortega (2006).

Language

We have found a wide diversity of elements of expression, representation and communication which conform the language. It can make the learning of the mathematical objects related to variation difficult.

Table 3: Frequency of language items

PUBLISHING HOUSE	LANGUAGE				
	Words/Expresions	Symbols	Formulas	Graphs	Tables
Anaya	12	6	10	15	61
Santillana	11	3	3	11	61
Bruño	13	4	3	5	35
McGrawH	14	5	4	8	25
Edelvives	18	6	5	2	69
Oxford	15	10	18	5	35
SM	11	3	4	7	19
Casals	12	5	6	17	47

Concepts. Definitions

Not all the definitions appear in all the books. We can see the frequency of the definitions found in Table 4.

Table 4: Frequency of the definitions analyzed

CONCEPT	DEFINITION	F
Dispersion (CD)	CD.R. Variation of the data around an average (Referential variation).	8
	CD.I. Variation between data (Intrinsic variation).	3
Range (CR)	CR.N. Range = max – min	8
	CR.T Length of interval where data are.	1
Interquartile Range (CIR)	CIR.D Difference between the third quartile and the first quartile.	1
	CIR.R Range of the middle 50% of the data.	1
Average Deviation (CAD)	CAD.N. Average of the absolute deviations of the data with respect to the mean of that data set.	5
	CAD.T Average of the distances of the data with respect to the mean of that data set.	2
Variance (CV)	CV.M. Average of the squared deviations of the observations from their mean.	*8*
	CV.T Average of the squared distances of the observations from their mean.	1
	CV.S2 The variance is the square of the standard deviation.	1
	CV.MM $\sigma^2 = a_2 - a_1^2$	*2*
Standard Deviation (CS)	CS.M Formula, $\left(\left(\sum_i (x_i - \bar{x})^2 \cdot n_i\right) / N\right)^{1/2}$	*3*
	CS.V2 The standard deviation is the square root of the variance.	*8*
	CS.MM Formula, $\sqrt{a_2 - a_1^2}$	*1*
Coefficient of Variation (CCV)	CCV.C The coefficient of variation of Pearson is the quotient between the standard deviation and the mean.	*2*
	CCV.100 $CV = (\sigma / \bar{x}) \cdot 100$ or CV = $(\sigma / \lvert\bar{x}\rvert) \cdot 100$	*1*

Properties

Three kinds of properties are considered: numerical, algebraic and statistical.

NP1. The ranges, the average deviation and the standard deviation are measured in the same units as the data. The variance is measured in square units of the data. The coefficient of variation is a-dimensional

NP2. All measure of variation is essentially nonnegative.

NP5. The standard deviation of n data is bounded both at the top and at the bottom.

AP4. If there is a linear transformation, the dispersion measures behave thus: $R(ax_i + b) = |a| \cdot R(x_i)$; $IR(ax_i + b) = |a| \cdot IR(x_i)$; $AM(ax_i + b) = |a| \cdot AM(x_i)$; $V(ax_i + b) = a^2 \cdot V(x_i)$; $S(ax_i + b) = |a| \cdot S(x_i)$

AP6. The formulas CV.MM and CS.MM are more used than CV.M and CS.M respectively.

SP1. The sum of the deviations with respect to the mean is zero.

SP2. König’s theorem (property of deviations): $\sum (x_i - \bar{x})^2 = \sum (x_i - a)^2 - n(\bar{x} - a)^2, \forall a \in R$

SP3. All the observations are equal if and only if the variation is null.

SP4. A greater value of the dispersion measures means greater dispersion.

SP5. In the calculation of the average deviation, variance, standard deviation and coefficient of variation take part all the values that the variable takes. On the contrary, in the calculation of the ranges not all the data take part.

SP8. The ranges are measures of the dispersion based on the order of the data.

SP9. The ranges measure the dispersion without reference to averages. The standard deviation and the variance measure the dispersion of the observations around the mean.

SP10. The coefficient of variation is useful to compare the variation of variables measured in different units or that take measured values in different magnitudes

SP15. For any distribution is true the inequality: $AM \leq S \leq \frac{R}{2}$.

SP16. The ranges, deviations and CV only are calculable for cuantitive variables.

Table 5: Frequency of actions analyzed and the publishers in which they appear

	PROPERTIES														
	NUMERICAL			ALGEBRAICAL		STATISTICAL									
	1	2	5	4	6	1	2	3	4	5	8	9	10	15	16
Anaya	3	0	0	2	2	1	3	0	6	0	0	0	2	0	1
Santillana	2	3	0	1	0	7	0	0	12	0	0	0	0	0	3
Bruño	2	0	0	0	0	1	0	1	11	1	0	0	0	0	0
McGraw	2	0	0	0	0	0	0	0	14	1	0	1	0	0	0
Edelvives	1	0	0	0	0	1	0	0	8	0	2	1	0	0	0
Oxford	0	2	1	4	1	1	0	3	12	0	0	0	0	2	0
SM	1	1	0	2	0	0	1	0	3	0	0	0	0	0	0
Casals	2	0	0	2	0	3	0	0	12	0	0	0	0	0	7
TOTAL	13	6	1	11	3	14	4	4	78	2	2	2	2	2	11

Arguments

The previous elements of meaning are leagued and related by means of arguments and reasonings in the textbooks analysed. The kinds of arguments that we have found, can be classified in the following way:

Table 6: Frequency of arguments found

	Anaya	Santillana	Bruño	McGraw	Edelvives	Oxford	SM	Casals	Total
Empirical	15	19	5	15	9	13	7	17	100
Deductive	2	2	5	1	1	2	2	7	22

DISCUSSION

In general, the following conclusions can be drawn:

- Only a few situations can be considered as true problems, looking for almost exclusively the dominion of skills of calculation.
- The techniques needed for the resolution of problems are mostly developed in numerical contexts. Graphs are rarely used to represent or interpret dispersion, and, in several textbooks, the use of formulas is avoided.
- All textbooks use words and expressions whose meaning can be confusing for students. For example, dispersion, difference, error, distance, homogeneity, heterogeneity, representativeness, concentration, etc.
- Confusion is introduced, sometimes deliberately, between referential and intrinsic dispersion, as well as between absolute and relative dispersion.
- There is no unanimity in the sequencing of contents dealing with dispersion. Some publishers include dispersion in 2nd E.S.O. (13 year old), whereas others postpone it to 4th E.S.O. (16 year old).
- Range, variance, and standard deviation are studied in all the analyzed textbooks. It’s not the same with interquartile range or the coefficient of variation.
- Finally, we have to emphasize the overuse of empiric reasoning, generally based upon examples chosen at random, impairing deductive ways of arguing.

Our results will be useful for the study, from the didactic point of view, that we will make in Compulsory Secondary Education. It can also be useful to plan the teaching of this topic, since we have detected many important elements of meaning. Finally, it can be useful for designing and planning of educative research in this topic.

ACKNOWLEDGEMENT

This research has been funded by “Programa de Promoción General del Conocimiento.” Dirección General de Investigación. Ministerio de Ciencia y Tecnología (Spain), project BSO2003-06331/PSCE.

REFERENCES

Cobo, B. and Batanero, C. (2004). Significado de la media en los libros de texto de secundaria. *Enseñanza de las Ciencias*, 22(1), 5-18.

Estepa, A. and Ortega, J. (2006). The meaning of the statistics variation in university text books. In A. Rossman and B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. Voorburg: The Netherlands: International Statistical Institute.

Font, V. (2004). La dimensión dual “personal/institucional” y el problema del encaje de los objetos personales del profesorado en la Teoría de las Funciones Semióticas. *First International Congress on Applications and Developments of the TFS*. Jaén. Spain.

Godino, J. D. (2003). *Teoría de las Funciones Semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática.* University of Granada. Spain. http://www.ugr.es/~jgodino/indice_eos.htm

Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.), *On the Shoulders of Giants: New Approaches to Numeracy*. Washington, D.C.: National Academy Press.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph and K. Carr (Eds.), People in mathematics education. *Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australasia Incorporated*, 1, 6-22. Waikato, New Zealand: Mathematics Education Research Group of Australasia.

Shaughnessy, J. M. and Ciancetta, M. (2001). Conflict between students’ personal theories and actual data: The spectre of variation. Paper presented at the Second Roundtable Conference on Research on Statistics Teaching and Learning. Armidale, New South Wales, Australia.

APPENDIX: TEXTBOOKS ANALYZED

[1] Colera, J., Gaztelu, I., Oliveira, M. J. and García, R. (2002). *Matemáticas 3ª ESO.* Anaya.

[2] Colera, J., Gaztelu, I., Oliveira, M. J. and Martínez, M. M. (2003)/ *Matemáticas 4ª ESO-A.* Anaya.

[3] Colera, J., Gaztelu, I., Oliveira, M. J., García, R. and Martínez, M. M. (2003). *Matemáticas 4ª ESO-B.* Anaya.

[4] Vizmanos, J. R. and Anzola, M. (1999). *Algoritmo 2000 (Matemáticas 3ºESO, 4ª ESO-A & 4º ESO-B).* Sm.

[5] Sánchez, J. L. and Vera, J. (1998). *Matemáticas 3ª ESO, 4º ESO-A & 4º ESO-B.* Oxford Ed.

[6] Lazcano, I. and Sanz, J. F. (1998). *Matemáticas 3º ESO & 4º ESO (Proyecto Adara).* Edelvives.

[7] Arias, J. M. and Pérez, S. A. (1995). *Matemáticas 3º ESO*. Casals.

[8] Arias, J. M., Carpintero, E. and Sanz, F. J. (1996). *Matemáticas 4º ESO-A & Matemáticas 4º ESO-B*. Casals.

[9] Alvarez, M., Ovejero, M. J., Bartomeu, C., Guiteras, J. M. Jané, A. and Moreno, J. (1998). *Matemáticas 3ª ESO.* McGraw-Hill.

[10] Amigo, C., Peña, P., Pérez, A., Rodríguez, A., Sivit, F., Asencio, M. J., and Vicente, E. (1997). *Matemáticas 4ª ESO-A & Matemáticas 4º ESO-B.* McGraw-Hill.

[11] Almodóvar, J. A., García, P., Gil, J. Vázquez, C., Santos, D. and Nortes, A. (1999). *Órbita 2000 (Matemáticas 3º ESO, 4º ESO-A & 4º ESO-B).* Santillana.

[12] Miñano, A. and Ródenas, J. A. (1998). *Matemáticas 3º ESO*. Bruño.