ATTENDING TO STUDENTS' THINKING ON BIVARIATE STATISTICAL DATA AT SECONDARY LEVEL: TWO TEACHERS' PEDAGOGICAL CONTENT KNOWLEDGE

Sandra Quintas[1], Rosa Tomás Ferreira[2], Hélia Oliveira[3]
[1]Research Unit of the Institute of Education of the University of Lisbon, Portugal
[2]University of Porto & CMUP, Portugal
[3]Institute of Education, University of Lisbon, Portugal
sandramquintas@gmail.com

In this study we compare and contrast aspects of the pedagogical content knowledge of two experienced secondary mathematics teachers when they teach bivariate data under the topic of statistics. In particular, we aim to understand how their knowledge of students is expressed in their instructional process. The data collected for this study was obtained from observation of some lessons taught by each teacher, semi-structured interviews and resources used in the lessons. This study highlights the need for these teachers to help students in reasoning more deeply about bivariate relationships, namely regarding aspects of structure and strength, model fitting, and the role of the linear regression model in predicting events. It also shows the importance of fostering students' comprehension of concepts related to correlation and regression when they work, in the classroom context, on tasks that incorporate bivariate data.

INTRODUCTION

Various authors state the complexity of teaching and learning bivariate data and relationships (e.g., Garfield & Ben-Zvi, 2008). The topic of bivariate data is often interpreted by teachers as a functional dependence of two variables in mathematics, neglecting data variation eventually around a possible trend. This situation may occur because of the lack of preparation of teachers during their undergraduate or teacher certification studies (Engel & Sedlmeier, 2011). More investigation is needed in the statistical field that inform about challenges to surpass in the process of teaching and learning (Shaughnessy, 2007). This paper focuses on describing and analyzing two teachers' PCK, having in account the specificity of the bivariate data teaching.

BACKGROUND

Teachers' professional knowledge goes beyond content knowledge (e.g., Ponte & Chapman, 2006). Shulman (1986), in his seminal work in this area, distinguished among three categories of content knowledge in teaching, namely, subject matter knowledge, pedagogical content knowledge and curricular knowledge. Since then, these categories have been reviewed and adjusted by other researchers in mathematics and statistics education (e.g., Burgess, 2011; Ball, Thames & Phelps, 2008). Ball et al. (2008) developed a framework named "mathematics knowledge for teaching" (MKT) composed by two major categories: subject matter knowledge (SMK) and pedagogical content knowledge (PCK). The latter incorporates three components: knowledge of content and teaching (KCT); knowledge of content and students (KCS); and knowledge of curriculum (KC). It is recognized that teachers' practices are privileged contexts that can reveal many aspects of teachers' PCK (Ponte & Chapman, 2006). Each PCK component in here considers the specificity of the statistics, as noted by several authors (Shaughnessy, 2007).

The KCT component refers to the capacity of sequencing the teaching and recognizing the advantages and disadvantages of using different representations and models; it also encloses the capacity of adapting statistical contents throughout different grade levels and of teaching based on students' active role in building mathematics and statistical knowledge (Garfield & Ben-Zvi, 2008); the KCS component refers to the knowledge of students' common misconceptions and of concepts and skills that they will struggle with or find easy; it also includes the perception of students' various levels of understanding of statistical content and their strategies in problem solving; the KC component includes the knowledge of curricular goals, instructional materials and the horizontal and vertical articulation of the various topics, as well as the comprehension of statistical big ideas (data, context, variation, distribution, representations, association and correlation, among others) expressed in the curriculum (Burrill & Biehler, 2011).

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

According to Garfield and Ben-Zvi (2008), in the teaching of statistics, reasoning about bivariate data goes beyond reasoning about a scatterplot; it involves ideas of structure and strength of a bivariate relationship, model fitting and residuals, and also the understanding of the usefulness of the regression model. Some authors pointed out some common misconceptions and errors with respect to bivariate relationships (e.g., Engel & Sedlmeier, 2011). For example, high correlation does not imply validity of a linear model, having the need of properly examining graphical representations of the bivariate data, since the correlation coefficient is a measure very affected by the presence of outliers. Also, the positive correlation between two variables translated as one increases, the other also increases, is not always a precise statement that characterizes this correlation, but instead, above average values of one variable correspond to above average values of the other. This detailed statistical knowledge can contribute for a better appreciation of the role of the variation and understanding of the coefficient correlation formula. Several authors declare that the type of tasks proposed by teachers and the way that they are worked in classroom influence the quality of students' learning (e.g., Garfield & Ben-Zvi, 2008). Activities that incorporate the establishment of connections between correlation coefficient values and scatterplots allow students to develop a better sense of different levels of covariation and an understanding of factors that may affect a higher or lower value of the correlation coefficient (Garfield & Ben-Zvi, 2008). It is important to propose tasks that include real data, elements about contexts and questions that make students appreciate the value of representations and reason about data (Curcio & Artz, 1996). Teachers should help students in gaining experience and comprehension about the modelling process in order to foster their understanding of correlation and regression (Engel & Sedlmeier, 2011). These activities may contribute to developing students' statistical reasoning.

METHODOLOGY

The study presented in this paper is part of a larger ongoing qualitative research that focuses on Portuguese secondary mathematics teachers, more concretely on teachers' pedagogical content knowledge in statistics. This paper refers to Estela and Lia, two case studies of the wider research. Estela and Lia are certificated mathematics teachers and Estela has earned a master degree in mathematics education. Each teacher has a career of at least 15 years and, taught one 10^{th} grade class with 25 students. The qualitative data that informed this paper were collected for each teacher using the following methods: (1) participant observation with audio and video recordings of three 90-minute classes where the theme bivariate data was taught; (2) documental collection of resources used by the teachers in these lessons, specially work sheets; and (3) two semi-structured interviews with audio recording. The data analysis was accomplished in a descriptive and interpretative way and guided by the PCK components. Estela and Lia followed the sequence of statistical contents suggested in the Portuguese secondary mathematics curriculum (DES, 2001): (1) generalities about statistics, population, sample, sampling; the statistical procedure: descriptive and inferential statistics; (2) organization of quantitative and qualitative data; and (3) introduction to the study of bivariate data (graphical and intuitive approach).

RESULTS

In general, with respect to KC, in the teaching of bivariate data Estela and Lia selected a set of tasks that targeted the fulfillment of current Portuguese secondary school program goals. In particular, the two teachers posed a few tasks that consisted of a set of scatterplots for making a correspondence with the respective linear correlation coefficients chosen from a set of given values. They also posed tasks that required analyzing the bivariate relationship between two quantitative variables, namely, by using the regression line. With respect to the KCT of the teachers, although their practices may be globally characterized by a teacher-centered approach, there are some differences between them. Estela sustains the introduction of contents and its discussion when the students are working on their tasks and, simultaneously, she poses questions that provide an orientation about what students should do, in a stimulating environment for student participation. Lia introduces the content through expositive moments, in which she conveys the necessary information, shows concepts and representations with the help of examples and gives explanations, which she intertwines with some questions. After the students' work on the proposed tasks, both teachers assure the correction of the answers given by the students collectively.

Estela and Lia suggested that students examine the *bivariate relationship of quantitative variables* represented in scatter plots taking into account the “signal” and the “strength” of the relationship. The “signal” was observed through the general tendency of data (growth/decrease) and its slope (positive/negative). Both teachers referred to the strength of the relationship as correlation stating that it could be “perfect, strong, moderate, weak or null”, and that it got stronger as data were closer to a line. Estela described also that this line should be “the line that best fits the data points of the scatterplot”. The positive correlation between two variables was always translated by both teachers as “one increasing when the other also increases”, even when they refer to distributions with several data that did not follow this trend. When Estela´s students addressed the task of associating scatter plots with correlation coefficients, the teacher also proposed the use of the graphical calculator as a first approach. The students who followed this suggestion had to insert all data represented in the scatter plot and the instructions to obtain the regression line equation and the correlation coefficient value, as the teacher indicated. Estela believed that such a recommendation was an opportunity to show the students the utility of the graphic calculator in exploring data. However, in class, her main goal for this strategy was to reach the coefficient correlation value. Both teachers reviewed these tasks with interaction with their students. In general, students did not show difficulties and denoted the right correspondences. In Estela’s case, students supported their answers reporting the “signal” and “strength” (strong, moderate, weak) of the relationship but they were never asked to justify them. In Lia’s case, students made the same kind of report and when explaining their choice for a lower coefficient they related it to the observation of a greater dispersion in the scatterplot.

With respect to the tasks that encompassed the use of the *regression line*, the two teachers incorporated real or fictitious quantitative data requiring examination of a bivariate relationship. During the process of solving one such task, aimed at introducing the regression line, Estela asked her students to obtain a line that they thought should be the best one fitting the set of bivariate data being analyzed. Then, she taught them how to determine the regression line using the graphic calculator. Some students reacted to the latter line being not equal to the one that was stated as the best that fits the data points of the scatterplot. As a student said: “But teacher, we didn’t get the same values [referring to the different slopes and intercepts of the two lines]”. Nonetheless, Estela argued that the regression line from the calculator was the most precise, and therefore, the one that they were looking for. She added that “the lesser the distance between the data points to the regression line, the stronger is the correlation” and there was a student that linked this idea to the fact that another student had shown her a calculator screen precisely with the distance of each data point to the regression line through squares. However, this comment was not explored by the teacher. Estela discussed with students some limitations of the linear regression model, for example, in estimating the evolution of the Portuguese population with data from 1864 to 2000, but without working specific examples. These situations show some missing opportunities of helping students to surpass difficulties and deepen their statistical knowledge, which may be influenced by Estela’s level of statistical understanding and by the way that she considers that statistical bivariate data fits in the curriculum. Lia defined the regression line in a similar way to Estela, as “the one that is closer from the scatterplot” without explaining the meaning of this expression, and gave the instructions for determining it on the calculator. As well, Lia underscored the utility of the regression line ($y=ax+b$) in predicting values for each variable (x or y) inside the intervals of variation of each one, apparently unknowing that a given regression line with independent variable x and dependent y, should not be used for predicting values of the independent variable. When solving a task that involved determining the regression line and performing predictions, Lia’s students in general did not have difficulties in finding the regression line on their graphic calculators. However, against the teacher’s expectations, for estimating the value of a variable (x or y) given the value of the other, they were unable to reach the result directly from the calculator, perhaps due to lack of experience in using this resource. Lia suggested they solve the question by hand with the line equation, but the students still had some difficulties finding x given y. In class, Lia never referred to the linear regression model and did not point out limitations of using the regression line for extrapolating results. These situations reveal Lia’s low priority in developing students’ statistical understanding on bivariate data, how she perceives that statistical bivariate fits in the curriculum and some deficiencies in her statistical knowledge.

DISCUSSION AND CONCLUSION

We observed in the analysis of the two teachers' PCK when they teach bivariate data in Statistics, some difficulties on how to proceed and support students in analyzing and interpreting the correlation coefficient and in reasoning about the regression line and the linear regression model. When viewing scatterplots, only linear relationships were explored even on graphs suggesting otherwise. Also, there was no analysis of the distribution form in terms of the existence of groups or *outliers*, nor how these might affect the correlation coefficient value. This experience may lead students to the idea that finding the correlation coefficient is enough for drawing conclusions about the existence of a linear relationship or the validity of the model, as Engel and Sedlmeier (2011) stated. Both teachers do not seem to have a strong experience in examining bivariate data represented by scatterplots, an activity that has potential to develop students' reasoning (Garfield & Ben-Zvi, 2008), which impacts negatively on their KCS. Although these classes were teacher-centered, the higher participation of Estela's students made more perceptible their difficulties and hurdles learning bivariate data. However, Estela did not always explore those difficulties in a way that promotes students' reasoning. For instance, although she was asked in class about the structure of the regression line and revealing some understanding about residual analysis, she didn't feel the need to deepen this knowledge of her students. On the other hand, Lia referring to the regression line has mentioned the minimum squares method but did not explain it. These situations, among others, may show the perception that both teachers have of the curriculum, their level of statistical understanding and also the level of demand on bivariate data that they require of students. This study highlights the importance of secondary school teachers developing their PCK in the teaching of bivariate data, as well as their knowledge in statistics.

ACKNOWLEDGEMENTS

Work funded by FCT, PORTUGAL, under the scope of the project *Developing statistical literacy: Student learning and teacher education,* grant # PTDC/CPE-CED/117933/2010.

REFERENCES

Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59(5),* 389-407.

Burgess, T. (2011). Teacher knowledge of and for statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school Mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 259-270). New York: Springer.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school Mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 57-70). New York: Springer.

Curcio, F., & Artzt, A. (1996). Assessing students ability to analyze data: Reaching beyond computation. *The Mathematics Teacher*, *89*, 668-673.

DES (2001). *Programa de Matemática A, 10.º ano*. Lisboa: ME, Departamento do Ensino Secundário. Available at: http://www.dgidc.min-edu.pt/ensinosecundario/

Engel, J., & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. In C. Batanero, G. Burril, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 247-258). New York: Springer.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching Practice*. Dordrecht, The Netherlands: Springer.

Ponte, J. P., & Chapman, O. (2006). Mathematics teachers's knowledge and practices. In A. Gutiérrez & P. Boeno (Eds), *Handbook of research on the psychology of mathematics education: Past, present, and future* (pp. 461-494). Roterdham: Sense.

Shaughnessy, J. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1009). Reston: NCTM.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15(2)*, 4-14.