VOCATIONAL TRAINING STUDENTS' READING LEVELS OF STATISTICAL GRAPHS

Pedro Arteaga, José Manuel Vigo, Carmen Batanero and José Miguel Contreras
University of Granada
parteaga@ugr.es

We analyse the responses given by 75 1st and 2nd grade vocational training students (hairdressing and aesthetics speciality), to several questions concerning interpretation of graphs (population pyramid and line graph). We propose a classification with five reading levels that combines those by Bertin and Curcio and evaluate the reading level reached by the students, using this new classification. In our results, few students reached the upper level of critically reading the graph and a considerable proportion did not reach the intermediate levels. We observed better performance in the 2nd grade students, who had studied some statistics the previous year. We conclude the need to reinforce the students' graphical competence to assist them to manage in the information society.

INTRODUCTION

Statistical graphs are pervasive in our society and can be used to communicate information efficiently, as a tool for data analysis; therefore, the construction and interpretation of statistical graphs is also an important part of statistical literacy (Gal y Murray, 2011; Kemp y Kissane, 2010). Moreover, the study of statistical graphs is included in compulsory education at the primary and secondary school levels. Finally, in their future professional work, the students can find statistical graphs similar to those included in our questionnaire, which has been taken from statistical studies carried out in the specialty of the students of our sample.

In this paper we present an exploratory study aimed at evaluating the reading level reached by Vocational Training students of the specialty of hairdressing and aesthetics before the formal teaching of the subject. Within Basic Vocational Training (MECD, 2014), mathematics does not appear as an independent subject, but is part of the Applied Sciences module. Statistical graphs are considered in the section "Graph Interpretation", where deterministic and statistical graphs are combined. In particular, the textbook used by the students in the sample (Brandi, 2015) includes histograms, bar graphs, frequency polygons, pie charts, pictograms and population pyramids. In the book students are asked to represent and read these graphs, translate data into tables or graphs, and identify the statistical graph that best fits the data. The aim of this research is to carry out an exploratory evaluation study to determine the graph reading levels of 1st and 2nd grade students in the specialty of Hairdressing and Aesthetics of Basic Vocational Training. Since the sample is small and intentional, we do not try to generalize the results to other students or context.

THEORETICAL FRAMEWORK

Many authors have described different levels of graphical understanding; Bertin (1967) proposed the following levels: *B1) Extracting data* or direct reading of the data represented on the graph. For example, in a bar graph, reading the frequency associated with a value of the variable; *B2) Extracting trends;* being able to perceive a relationship between two data subsets that can be defined a priori or visually in the graph. For example, visually determining the mode of a distribution in a bar graph; *B3). Analysing the data structure:* comparing trends or clusters and making predictions. An example is analysing the differences in mean and range of two distributions in an attached bar graph.

Curcio (1989) proposed a related classification with the following levels: *C1) Reading the data* (literal reading of the graph without comparing the information contained in it), C2) *Reading between the data* (interpreting and integrating the data in the graph), *C3) Reading beyond the data* (making predictions and inferences from the data to information that is not directly reflected in the graph); this classification was expanded by Shaughnessy, Garfield and Greer (1996), with the level *C4) Reading behind the data*, which consists of judging the method of data collection, and assessing the data validity and reliability, as well as the possible generalization of findings. In our work, to facilitate the codifying of student's' answers we condense the two previous hierarchies

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

into one and propose the following reading levels:

- *N0: Incorrect reading:* The student does not answer the question or gives an incorrect reply.
- *N1: Reading the data* (levels B1 or C1). The student only reads simple data of the graph.
- *N2: Extracting trends in a single distribution (*levels B2 and C2): The student compares data or performs calculations with them within a single data distribution.
- *N3: Extracting the structure in a multiple data representation* (levels B3 and C2): Comparing the trends of two or more datasets in graphs representing two or more distributions.
- *N4: Reading beyond the data* (level C3): Predicting a value that is not in the graph, i.e., interpolating or extrapolating the data.
- *N5: Reading behind the data* (level C4): Giving a critical interpretation of the contents of a graph. The relevance of this critical level for all citizens who should acquire insight into how information is presented in the graphs and to be able to ask critical questions about how this information and the conclusions that are drawn were emphasized by Aoyama (2007).

Research describing reading levels of high school students is scarce. One such research is that by Fernandes and Morais (2011) with 108 students of 9th grade, using bar graphs, pie charts and line graphs, which were the most difficult graphs (25.3% correct answers, with 45.3% correct response in the other graphs). Only 24% of students answered Level 2 questions, and 33% Level 3 questions in Curcio's (1989) classification, while Level 1 questions were answered by 68% of students. Pagan and Magina (2011) conducted a study with 105 students of 9th grade, based on the application of a pre-test, a classroom intervention and a post test, which includes activities of reading graphs. The authors investigated the Curcio reading levels, 67% of students reached level 1, 42% level 2 and 18.7% level 3. Carvalho, Campos and Monteiro (2011) analysed the direct and inverse reading of line graphs in English students (84 students from 7th to 9th grades); 74.1% successfully did the direct reading and 37.7% the reverse, improving the results with age.

Our work is based on this previous research, but uses the same graph to raise different questions in which students can reach level 4 of Curcio (level 5 in our classification). We also propose the classification of levels that serves to combine the categories of Bertin and Curcio.

METHODOLOGY

This work was carried out in a sample of 47 students, of whom only one was a boy, of two different grades (two first-grade groups with a total of 29 students and one second-grade group with 18 students) attending Basic Vocational Training in the specialty of hairdressing and aesthetics in an educational centre of Ceuta, Spain. The socioeconomic level of the students was varied and highly differentiated, both in social class, religion and previous training; the approximate age was 15-16 years old. We proposed five items to the students where they had to interpret different statistical graphs taken from the Internet. In this work we analyse only two of these items (Figures 1 and 2). Item 1 is taken from a study on consumption in hairdressing and represents three series of data: monthly average expenditure on hairdressing of men, women and in both genders. Item 2 is taken from a study of dermatology care demand of men and women at different ages. We expect the following responses to the questions shown in Item 1 (Figure 1):

a) The student has to observe the variation throughout time of hairdressing expenditure of men and women, which tends to increase. They should also see that the hairdressing expenditure is always higher in women. So, they have to determine and compare the trend in the two series. It is a question of level N3 in our classification (level B3 in Bertin and C2 in Curcio).
b) The second question asks to predict data that is not in the graph. To answer the question the student should realise that in men there is an increase of 7 euros (1.5 per year) and in women about 10 (2 euros per year) and use this estimation to find the response. Another reasonable answer is obtained by increasing the estimated amount per year and adding the result to the last data shown in the graph. The increase in average expenditure was higher the last year represented in the graph than the previous years, in both men and women, so it can be assumed that the prices of services rose. If the students respond in this way they reach level N4 in our classification (Level C3 in Curcio and not considered by Bertin).
c) Although the third question is similar to the previous one, in fact we expected that some students could observe that it is difficult to give the prediction with so much time spent without

current data (7 years of difference), since the price could have changed too much and even decrease (by a greater offer of hairdressers or lower consumption, due to the economic crisis of the past years). If the students respond in this way they reached level N5 in our classification (Level C4 in Curcio´s classification).

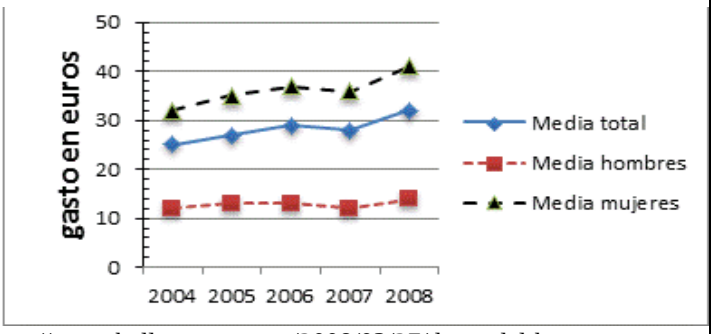
Item 1. The following graph displays the evolution of the average monthly expenditure in euros in hairdressing in the period 2004-2008.

a. Was the evolution of the monthly average expenditure in hairdressing in men and women similar throughout 2004-2008?

b. Using the data displayed in the graph, what do you think was the monthly average expenditure on hairdressing in 2009?

c. What do you think was the monthly average expenditure on hairdressing in 2015?

Average monthly expenditure in hairdressing



http://www.bellezapura.com/2008/03/27/datos-del-barometro-cosmobelleza-2008/

Figure 1. Item 1 presented to the students

Item 2. The following graph displays the healthcare demand for dermatological care in men and women at different ages.

The data represent the number of people who attended in one year a hospital dermatology service.

a. What differences do you observe between the men and women´s demand for dermatological assistance?

b. In what age range is the demand for care more likely in men and women? In what range is there more variation?

c. Why do you think there are such differences?

Dermatology care demand

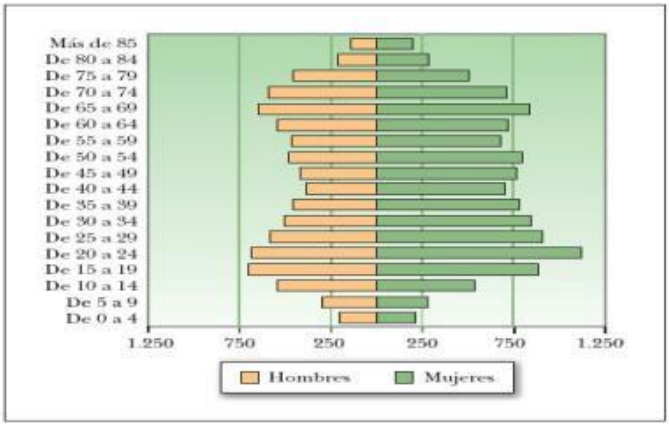


Figure 2. Item 2 presented to the students

DETAILED RESULTS IN ITEM 1

Once the questionnaires were collected, we assigned each student´s responses a numerical value 1 to 5, according to the reading level reached in his/her answer and 0 if the students did not answer or made an incorrect reading. Below we describe the answers corresponding to each task in item 1, and include an example to clarify the way in which the responses have been codified.

First Task. Evolution of average expenditure in men and women

The response to task a) in item 1 was assigned a level N0 if the student was unable to correctly read the data. In level N1, the student read some data in the graphs, but did not answer the question posed; for example, the response was “Not” with no justification. In level N2, the student perceives the trend for only one group (men or women), and is unable to compare both tendencies, for example, the following student observes that the monthly average expenditure in women is higher than in men, but makes no comment on the trend over time: "No, women spent more ". In level N3 the student provides a correct answer, by clearly observing the variation in both data sets

throughout time and observing that the expenditure is higher in women than in men. For example: "Yes, the expenditure increases equally in both groups, although women always spend more".

Second Task. Extrapolating a close value

In Task b) the student was assigned level N0 when providing no response or making an incorrect reading. Level N1 corresponds to a student who correctly reads some data in the graph, but does not answer the question; for example, "in 2008 the average expenditure was over 30 €". The student is assigned level N2 when he or she correctly reads the data, and identifies the trend of the data sets. However, the predicted value is far away from what was expected and the student does not justify this prediction. For example: "Around 45 to 50 €".

N4: If the student responds correctly to the question posed, being able to compare the trends of two data sets and provides a possible value, taking into account the graph structure in a reasoned way. In the following example, we observe that the student gives a numerical value, which although does not identify that men's expenditure increases 7 euros (1.5 per year) and women 10 euros (2 euros per year), he gives a reasonable estimation: "40 euros". We do not consider level N3 in this task because the students are not asked to compare two trends.

Third Task. Extrapolating to a distant value

In Task c) the lack of response was classified in level N0. The student was assigned a level N1 when he was able to interpret the graph, but did not provide an estimation. Level N3 corresponds to students correctly identifying or comparing the trends of the two data sets: "Women will spend more than men".

In level N4 the student responds to the question by giving a roughly correct value, but does not reason that it is difficult to predict without more data, i.e. the student does not reach the critical level because, although able to determine the graph tendency and extrapolate, he or she does not reason correctly. In the following example, the student gives an acceptable numerical data, but does not provide any reasoning: "46". In level N5 the student is able to determine the monotony of the graph and predicts a reasonable value; in addition, he should highlight the difficulties of making this estimate, due to the number of years in which data are not available. This level has not been reached by any students.

Table 1. Percentage of students achieving different reading levels in Item 1

	Grade 1 (n=29)				Grade 2 (n=18)			
Part	N1	N2	N3	N4	N1	N2	N3	N4
A	51.7	17.2	31.1		22.2	44.4	33.3	
B	27.6	44.8		27.6	44.4	22.2		33.3
C	44.8		6.9	48.3	22.2			77.8

A synthesis of the results in item 1 is shown in Table 2. We can see that in Task a) (evolution of the trend) the maximum level was N3, and only a third of the students reached this level. For the first-grade group, more than 50% stay at N0, while most students in the second group reached N2, where students establish a correct reading of the data, and are able to establish comparisons, but are unable to compare the two distributions. In Task b) (short-term prediction), we observe similar results, although some students reached the N4 level, maximum in this question. In the second grade, almost a third of the group reached the maximum level, while in the first grade, the percentage was somewhat lower. In part c) (long-term prediction) we expected the students to reach the N5 level, but none reached that level. The percentage situated at level N4 was almost 50% in the first grade and more than 75% in the second grade. However, the number of students who did not respond or was unable to carry out a simple reading of the graph was very high. Previous research only takes into account Curcio level 3 (our level 4 (N4)), reaching 33% in Fernandes and Morais (2011), 18.7% in Pagan and Magina (2011). Our results are similar in the second task but better in the third.

SUMMARY OF RESULTS IN ITEM 2

A similar analysis was carried out for the responses to item 2 and provided the results

presented in Table 2. In part a) more than half the respondents in both grades reached level N3, which was the maximum possible level in the question. These students identified the structure of the graph and made an appropriate analysis of the same. With regards to task b), the majority of students in Grade 1 reached Level 1, and more than 90% did not attain the desired level (N3). Students in Grade 2 performed better, although there is division of students, half of which reached level 1 while the remaining only achieved Level 3. Again, in task c) students in Grade 2 performed much better than those in Grade 1 and the majority of the same reached level N5. It is also worth mentioning the high percentage of students in both courses who did not ask the question or provided an incorrect response (level N0) in the different parts of this item. Results from grade 2 students are better than those in Fernandes and Morais (2011) and Pagan and Magina (2011).

Table 2. Percentage of students achieving different reading levels in Item 2

	Grade 1 (n=29)				Grade 2 (n=18)			
Part	N0	N1	N3	N5	N0	N1	N3	N5
a	24.1	17.2	58.6		33.3		66.7	
b	13.8	79.3	6.9			44.4	55.6	
c	37.9	17.2	37.9	6.9			22.2	77.8

Although the limited sample size makes generalization difficult, apparently the tasks proposed in item 2 were easier for the students in both groups that those included in item 1 (Figure 1) in agreement with Fernandes and Morais (2011) who found line graphs difficult for students. However no students stayed at level N0 in item 1. It is worth noticing that in all the tasks students in Grade 2 reached higher levels than their colleagues, which we attribute to the work with statistical graphs included in the teaching received by these students.

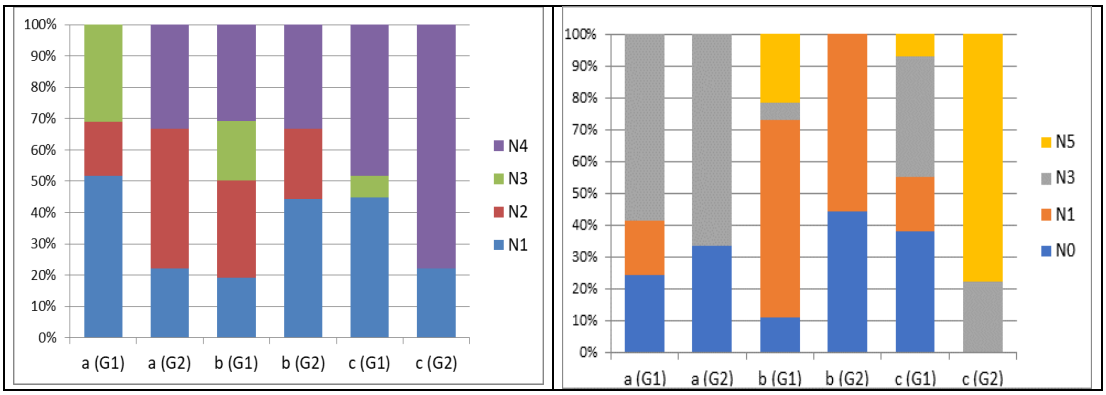


Figure 3. Comparing G1 and G2 reading levels in the different tasks in items 1 (left) and 2 (right)

DISCUSSION AND IMPLICATION FOR TEACHING

The graphs used in this research are recommended even in primary school and appear frequently in textbooks at this educational level, as shown by Diaz-Levicoy, Batanero and Arteaga (2015). Therefore, students should be familiar with these representations, at least for a literal reading. However, in our study many students did not achieve the first level of reading N1 that supposes only a literal reading of isolated elements of the graph in item 2.

We observed better results in the interpretation of statistical graphs by the students of the second grade of Vocational training in our sample, which is possibly due, on the one hand, to the degree of maturity of students in a more advanced grade and on the other hand, to the education that these students received at the end of the 1st grade. However, the low percentage of students who reached the maximum level in each question is a matter of concern. In addition, the following errors have been often observed in the responses of the students:

- Confusing the question posed, which suggests a poor reading comprehension competence. In this sense we agree with Gal and Murray (2011) that reading competence is a basic component of statistical literacy,
- Interpreting means represented in the graph as a simple value of a data, this error was also

found in Carvalho, Campos, and Monteiro (2011).

- Not identifying that the average expenditure in item 1 is computed for a whole year, and interpreting instead that the quantity only refers to a specific instant of time.
- Not being able to compare two distributions in one of the items; although the students can analyse each separate distribution they are unable to conclude from the differences between these distributions.
- Not identifying the growing trend in the distributions displayed in item 1.

The teacher should pay attention to the aforementioned errors and help students to achieve a sufficient level of statistical literacy that allows them to make a critical reading of the graphs they find in the media and in their professional life (Kemp & Kissane, 2010). Consequently, more time should be devoted to the study of statistics graphs and to performing interpretative activities. The reading of the graphs seems, at first sight, a simple activity and is taken for granted, and therefore little time is devoted to its teaching; the teaching time is instead spent on other subjects, for example, measures of central position or spread. Our research and those cited in the literature review show that reading a graph is difficult and we must reinforce this ability in students to help them achieve a good level of statistical literacy.

Acknowledgements: Project EDU2016-74848-P and Group FQM126 (Junta de Andalucía).

REFERENCES

Aoyama, K. (2007). Investigating a hierarchy of students' interpretations of graphs. *IEJME-Mathematics Education 2*(3), 298-318.

Bertin, J. (1967). *Semiologie graphique* (Graphic semiology). Paris: Gauthier-Villars.

Brandi, A. (2015). Matemáticas 1. Módulo de ciencias aplicadas. (Mathematics I. Applied sciences module) Madrid: Santillana.

Carvalho, C., Campos, T. M, & Monteiro, C. (2011). Aspectos visuais e conceituais nas interpretações de gráficos de linhas por estudantes. *Boletim de Educação Matemática, 24*(40), 679-700.

Curcio, F. R. (1989). *Developing graph comprehension.* Reston, VA: N.C.T.M.

Díaz-Levicoy, D., Arteaga, P. y Batanero, C. (2015). Gráficos estadísticos y niveles de lectura propuestos en textos chilenos de Educación Primaria (Statistical graphs and reading levels in chilean school textbooks). In C. Fernández, M. Molina y N. Planas (Eds.), *Investigación en Educación Matemática XIX* (pp. 229-238). Alicante: SEIEM.

Fernandes, J. A., & Morais, P. C. (2011). Leitura e interpretação de gráficos estatísticos por alunos do 9º ano de escolaridade (9º grade students' reading and interpretatioin of statistical graphs). *Educação Matemática Pesquisa, 13*(1), 95-115.

Gal, I., & Murray, S. T. (2011). Responding to diversity in users' statistical literacy and information needs: Institutional and educational implications. *Statistical Journal of the International Association for Official Statistics, 27*(3-4), 185-195.

Kemp, M., & Kissane, B. (2010) A five step framework for interpreting tables and graphs in their contexts. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. Ljubljana, Slovenia: International Statistical Institute. Disponible en http://researchrepository.murdoch.edu.au/6240/.

MECD (2014). Real Decreto 127/2014, de 28 de febrero, por el que se regulan aspectos específicos de la Formación Profesional Básica de las enseñanzas de formación profesional del sistema educativo, se aprueban catorce títulos profesionales básicos, se fijan sus currículos básicos y se modifica el Real Decreto 1850/2009, de 4 de diciembre, sobre expedición de títulos académicos y profesionales correspondientes a las enseñanzas establecidas en la Ley Orgánica 2/2006, de 3 de mayo, de Educación. Madrid: Author.

Pagan, A., & Magina, S. (2012). O ensino de estatística na educação básica com foco na interdisciplinaridade: um estudo comparativo (Teaching of statistics in basic education with focus in interdisciplinarity; a comparative Study). *Revista Brasileira de Estudos Pedagógicos, 92*, 232- 240.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), International handbook of mathematics education (pp. 205-237). Dordrecht, The Netherlands: Kluwer Academic Publishers.