INTRODUCING THE NORMAL DISTRIBUTION BY FOLLOWING A TEACHING APPROACH INSPIRED BY HISTORY: AN EXAMPLE FOR CLASSROOM IMPLEMENTATION IN ENGINEERING EDUCATION

Mónica Blanco

Marta Ginovart

Department of Applied Mathematics III

Technical University of Catalonia, SPAIN

Abstract: *Probability and random variables turn out to be an obstacle in the teaching-learning process, partly due to the conceptual difficulties inherent in the topic. To help students get over this drawback, a unit on "Probability and Random Variables" was designed following the guidelines of the European Higher Education Area and subsequently put into practice at an engineering school. This paper focuses on the design, implementation and assessment of a specific activity of this unit concerning the introduction of the normal probability curve from a teaching-learning approach inspired by history. To this purpose a historical module on the normal curve elaborated by Katz and Michalowicz (2005) was adapted to develop different aspects of the topic.*

Keywords: probability, normal distribution, European Higher Education Area, teaching-learning materials on history of mathematics.

INTRODUCTION

Teaching probability and random variables turn out to be essential for the introducing of statistical inference in any undergraduate course in basic statistics. Statistics is one of the compulsory undergraduate subjects included in the syllabus of any engineering school. This subject, as developed at the School of Agricultural Engineering of Barcelona (ESAB) of the Technical University of Catalonia (Spain), primarily encompasses Data Analysis and Basic Statistical Inference. We believe that the very nature of the subject calls for special consideration in the teaching of the subject, especially with regard to the new European Higher Education Area (EHEA). Besides, the essentially biological profile of the ESAB seems to weaken interest in mathematical domains.

From our experience in teaching statistics at different engineering schools, we are well aware that probability and random variables represent a rather overwhelming obstacle for students, due to the conceptual difficulties inherent in the topic. To help students get over this drawback, a unit on "Probability and Random Variables" was designed following the guidelines of the EHEA. Subsequently, this unit was put into practice at the ESAB. Throughout the module, the teaching-learning process was assessed using several evaluation techniques so as to analyse the learning outcome (Blanco & Ginovart, 2008). This paper focuses on the design, implementation and assessment of a specific activity of this unit concerning the introduction of the normal probability curve and some related aspects from a historical point of view.

Mathematical and statistical topics have been traditionally taught in a deductively oriented manner, presented as a cumulative set of "polished" products. Through a collection of axioms, theorems and proofs, the student is asked to become acquainted with and competent in handling the symbols and the logical syntax of theories, logical clarity being sufficient for the understanding of the subject. As a result, the traditional teaching of mathematics tends to overlook the mistakes made, the doubts and misconceptions raised when doing mathematics, detaching problems from their context of origin. However, since the construction of meaning is only fulfilled by linking old and new knowledge, the learning of mathematics, in general, and statistics, in particular, lies in the understanding of the motivations for problems and questions. In this respect, integrating the history of mathematics in education represents a means to reflect on the immediate needs of society from which the mathematical problems emerged, providing insights into the process of constructing mathematics (Tzanakis & Arcavi, 2000; Swetz et al., 1995).

How to introduce a historical dimension in our unit on probability and random variables turned out to be a challenge to our "standard" teaching activity, all the more so because first we had to determine which role history would play in the unit. Of the three different ways suggested by Tzanakis & Arcavi (2000) to integrate history in the learning of mathematics, the one that seemed to serve our purpose best was to follow a teaching-learning approach inspired by history. In the context of this paper history was integrated implicitly, since the main aim was to understand mathematics (statistics, in particular) in its modern form, bearing in mind, throughout the teaching process, those "concepts, methods and notations that appear later than the topic under consideration" (Tzanakis & Arcavi, 2000, p. 210). Accordingly, after having selected a historical module on the normal curve elaborated by Katz and Michalowicz (2005, pp. 40-57), we adapted it to develop different aspects of the topic. The aims of the activity were to:

Aim 1.- Show motivation for the topic.

Aim 2.- Show interrelation between mathematical domains, on the one hand, and mathematical and non-mathematical domains, on the other.

Aim 3.- Compare modern "polished" results with earlier results.

Aim 4.- Produce a source of problems not artificially designed for the purpose.

Aim 5.- Develop "personal" skills in a broader educational sense.

These aims are explicitly connected with the ones described by Tzanakis & Arcavi (2000, §§7.2. (a) and 7.2. (c1), pp. 204-206).

THE NORMAL DISTRIBUTION: AN INTRODUCTION INSPIRED BY HISTORY

Right at the beginning of the course our students are informed about the specified learning outcomes, classified according to Bloom's taxonomy (Bloom, 1956) into: Knowledge, Comprehension and Application. The learning outcomes regarding the normal distribution have been articulated as follows:

Table 1. Learning outcomes regarding the normal distribution.
After attending the course the student will be able to:

a) Define and recognize the normal (or Gaussian) distribution, as well as the standard normal distribution.	[Knowledge]
b) Convert an arbitrary normal distribution to a standard normal distribution.	[Comprehension]
c) Calculate probabilities of events when a normal distribution is involved, using the table of the standard normal distribution.	[Comprehension]
d) Describe the empirical rule 68-95-99.7.	[Comprehension]
e) Apply the rule 68-95-99.7 to assess whether a data set is normally (or approximately normally) distributed.	[Application]
f) Estimate the approximation of the normal distribution to the binomial distribution.	[Application]

To adapt the historical module it was first necessary to frame the activity within well-defined boundaries (Katz & Michalowicz, 2005). Therefore, we started selecting and later reflecting on some questions suggested by Pengelley (2002) for assessing historical material: (a) What is the purpose of studying the material? (b) How does it fit in with the curriculum? (c) Are there appropriate exercises, with an appropriate difficulty level and well chosen to demonstrate concepts? (d) Will it motivate students? (e) Will it help with something students have trouble with? Since the activity described in this paper was directed towards the learning outcomes mentioned above (see Table 1), question (b) was explicitly involved.

To show the original motivation for the topic of the normal distribution, the activity emphasized interrelation between statistics and health and social sciences, hence covering Aims 1, 2 and 4. Although the topic had already been introduced in the classroom, the teaching-learning process was able to benefit from the study of non-artificially designed problems. From Katz and Michalowicz's module we elaborated the material for the activity combining information about the historical development of the normal curve with some "appropriate" questions. There were no accompanying answer sheets as the activity was designed to be worked out in a two-hour computer lab session, individually or in pairs. Most of the students worked individually, whereas only few computers were shared by two students working together. The teacher acted as a consultant during the session. Students managed the time given over to every section of the activity themselves, according to their individual needs and skills. If they could not accomplish their work in the computer lab, they had the possibility to do it as homework. It is worth pointing out that the questions were chosen not only to assess understanding of the information provided, but also to bring out the connection with other mathematical domains. Hence, students were asked to prove expressions and formulae, to use a spreadsheet to carry out elementary probability calculations and to represent data, and to investigate supplementary aspects regarding the contents of the activity. All these aspects were planned in order to cover Aims 3 and 5.

In connection with question (a) stated above, this activity attempts to introduce the normal probability distribution in its original context, and to help students to get acquainted with basic calculations involving the normal curve. The first section of the activity shows how De Moivre (1667-1754) obtained his discovery of the empirical rule 68-95-99.7. The second section gathers the discussion on the error curve in which Laplace (1749-1827) and Gauss (1777-1855) were involved. How Quetelet (1796-1874) calculated the table of the normal distribution from the approximation of the normal distribution by the binomial distribution is the target of the third section. To close the activity, the fourth section is centered on the first uses of the normal distribution in the real world, namely: i) analysis of the chest circumference of 5732 Scottish soldiers; ii) analysis of the heights of French conscripts to assess the normality of the distribution, revealing a significant figure of men who illegally avoided recruitment.

We interspersed the text with seven leading questions related to the topics discussed, conveniently placed after a specific topic, and not on a separate sheet at the end. *Questions 1*, *4*, *6* and *7* were directly inspired by the ones suggested by Katz and Michalowicz (2005) on pages 46, 55, 56 and 57, respectively. The rest were stated by us, to ensure that a particular point was fully understood. The questions were conveniently placed after a specific topic or a related result. The following paragraphs briefly describe each question, drawing attention to the educational aims served by each one.

Question 1: In an experiment in which 100 fair coins are flipped, about how many heads would you expect to see? What is the corresponding standard deviation? Find the limits (lower and upper) for the number of heads we would get 68%, 95% and 99.7% of the times.

This first question deals with direct manipulation of a binomial distribution, followed by a first encounter with the connection between the normal and the binomial distributions. This was intended to help students “warm up” by stating a link between the activity and a topic they had already learned in the classroom, thus relating to Aim 1.

Questions 2 through *4* are connected with Quetelet’s calculation of a symmetric binomial distribution. He considered the experiment of drawing 999 balls from an urn containing a large number of balls, half of which were white, and half black.

Question 2: Prove Quetelet's shortened procedure for the calculation of relative probabilities: $P(X = n+1) = \frac{999-n}{n+1} \cdot P(X = n)$, where $P(X = n)$ represents the probability of drawing *n* black balls from the urn. Setting the value of $P(X = 500)$ to be 1, calculate the relative probabilities $P(X = 501)$ and $P(X = 502)$.

Students had to deduce this recursive formula from the probability function of the binomial distribution. This question was inserted to show the interrelation between mathematical domains, namely, probability and recursive proofs (Aim 2). In this case the interest lies in how to evaluate mathematical arguments and proofs, and to select and use diverse types of reasoning and methods of proof as appropriate (Ellington, 1998). Given that students often meet difficulties in proving recursive formulae, this exercise seems to be consistent with questions (c) and (e) suggested above.

Question 3: Using an Excel worksheet recalculate column A of Quetelet's table for the values 500 to 579 and graph the corresponding curve.

To get a deeper knowledge of the binomial-normal link, students were here asked to use a spreadsheet, in particular, the spreadsheet program Microsoft Excel. Since the activity was developed in the context of computer practical sessions, students had computers at their disposal. The computer practicals offer students the possibility to be actively engaged in the learning process, as well as to apply the concepts learnt to the prospective working practice. Since this topic turns out to be a usual source of difficulty, this exercise connects again with question (e). Besides, it helps not only to compare modern results with earlier ones, but also to develop "personal" skills such as how to manipulate a spreadsheet. Therefore, this exercise focuses on Aims 3 and 5.

Question 4: A discrete variable can be approximated by a continuous variable considering the following estimation: $P(x = k)_{discrete} \approx P(k - 0.5 \le x \le k + 0.5)_{continuous}$. For instance, $P(x = 500)_{binomial} \approx P(499.5 \le x \le 500.5)_{normal}$. Using this information, recalculate the first four values in column A using a modern table of the normal distribution. It can be assumed that the results of drawing balls out of the urn are normally distributed with mean of the number of black balls equal to 500 and standard deviation equal to $\frac{1}{2}\sqrt{999} \approx 15.8$. Compare these results with Quetelet's binomial table.

Understanding why we do things the way we do, and how mathematical concepts, terms and symbols arose, plays a relevant role in grasping the topic (Ellington, 1998). This question allowed the students to compare a modern table of the normal curve with the earliest table. Thus Aim 3 is again involved in the proposed activity.

Finally, *Questions 5, 6* and *7* concern some real world applications of the normal distribution.

Question 5: Read carefully Quetelet's procedure for determining whether the chest circumferences of the Scottish soldiers were normally distributed. Write down those points you do not understand completely.

Question 6: From the results in the example of the heights of French conscripts, discuss how Quetelet concluded there had been a fraud.

From the reading and through understanding of the example on the chest circumferences (*Question 5*) students were to draw conclusions in the case of the heights of French conscripts (*Question 6*). However, as we will see in the following section, since Quetelet's procedure proved to be difficult to understand, only a few students managed to answer *Question 6* correctly.

Questions 4, *5* and *6* contribute to Aim 3 in that they help to compare historical results with modern "polished" ones. Likewise, Aim 4 could be achieved, since these questions convey the idea that probabilistic tools represent a means to solve real-world problems, rather than just artificial designed exercises, framed in a theoretical context. By and large, this set of questions also fosters the practice of reading comprehension skills (Aim 5).

Question 7: On the Internet, browse for information on Galton's machine. What was the relationship between the inventor Francis Galton (1822-1911) and Charles Darwin (1809-1882)?

The intend of this last question was to help develop some "personal" skills, in a broader educational sense, such as reading, summarising, writing and documenting (Aim 5). Additionally, it was interesting to point out the interrelation between mathematical and non-mathematical domains, namely, between statistics and the theory of evolution put forward by Darwin (Aim 2). A fundamental part of this question involves the writing component and documenting. The incorporation of a writing component in statistics courses has been encouraged in recent years by Radke-Sharpe (1991) and Garfield (1994). Writing helps students to think about the assumptions behind statistical, graphical or instrumental procedures, to formulate these assumptions verbally, and to critically examine the suitability of a particular procedure based on its assumptions. The inclusion of documenting (i.e. browsing the Internet) facilitates student reading, understanding and summarizing from different sources. In short, reading, writing and documenting are tools that will serve students well in their future scientific or academic writing. Encouraging students to put concepts such as these into words will strengthen their understanding of those concepts.

ASSESSMENT OF THE TEACHING-LEARNING PROCESS

Among the questions mentioned above for assessing historical material, Pengelley (2002) suggests considering whether it will motivate students (question (d)). Though not the only source of feedback, student ratings provide an excellent guide for designing the teaching-learning process and, in particular, for assessing their motivation. Therefore, at the end of the activity students were asked to rate the activity thus:

(1) Very good, (2) Good, (3) Satisfactory, (4) Poor, and (5) Very poor.

Figure 1 shows the results of this survey. Of the 60 students who took part in the activity, half of them regarded it positively (22 satisfactory, 6 good, 1 very good), whereas the other half rated it as poor.

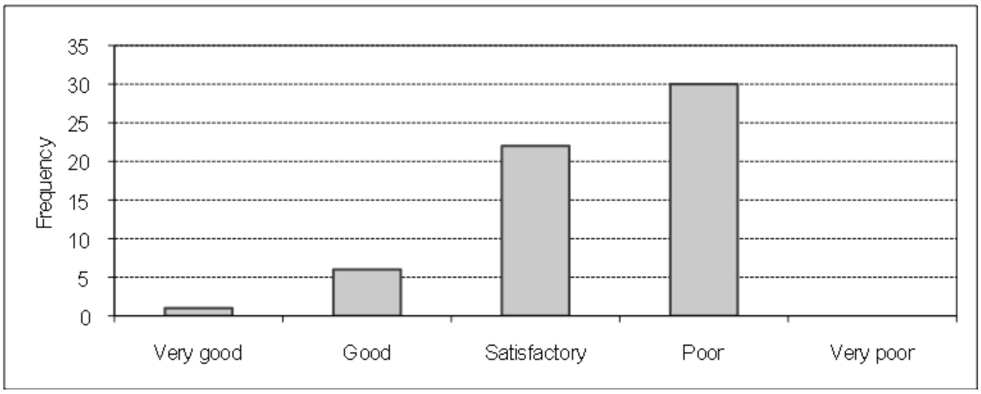

Figure 1. Student ratings on the activity.

Another aspect suggested by Pengelley (2002) for assessing historical material concerned the suitability of the degree of difficulty (question (c)). In order to determine whether the activity was appropriately difficult, we analysed in detail a random sample of size 20 drawn from the students who had handed in their answers. Every question (except *Question 5*) was marked with either Non-Answered, Poor, Fair or Good. From the graphics of Figure 2 regarding the assessment of the questions, it is clear that *Questions 1* through *4* are most frequently marked as "Good". Surprisingly, all the students answered *Questions 1* and *2*, whereas the ratio of "Non-Answered" in *Question 6* exceeded the rest of marked ratios. As for *Question 7*, most of the students got "Fair". This was partly due to the fact that students merely copied the information from the Internet and pasted it on their worksheets, thus showing no interest in summarising the information in their own words.

Relating to *Question 5*, from the comments given by our students we gathered that the construction of the table proved to be, in general terms, rather cumbersome.

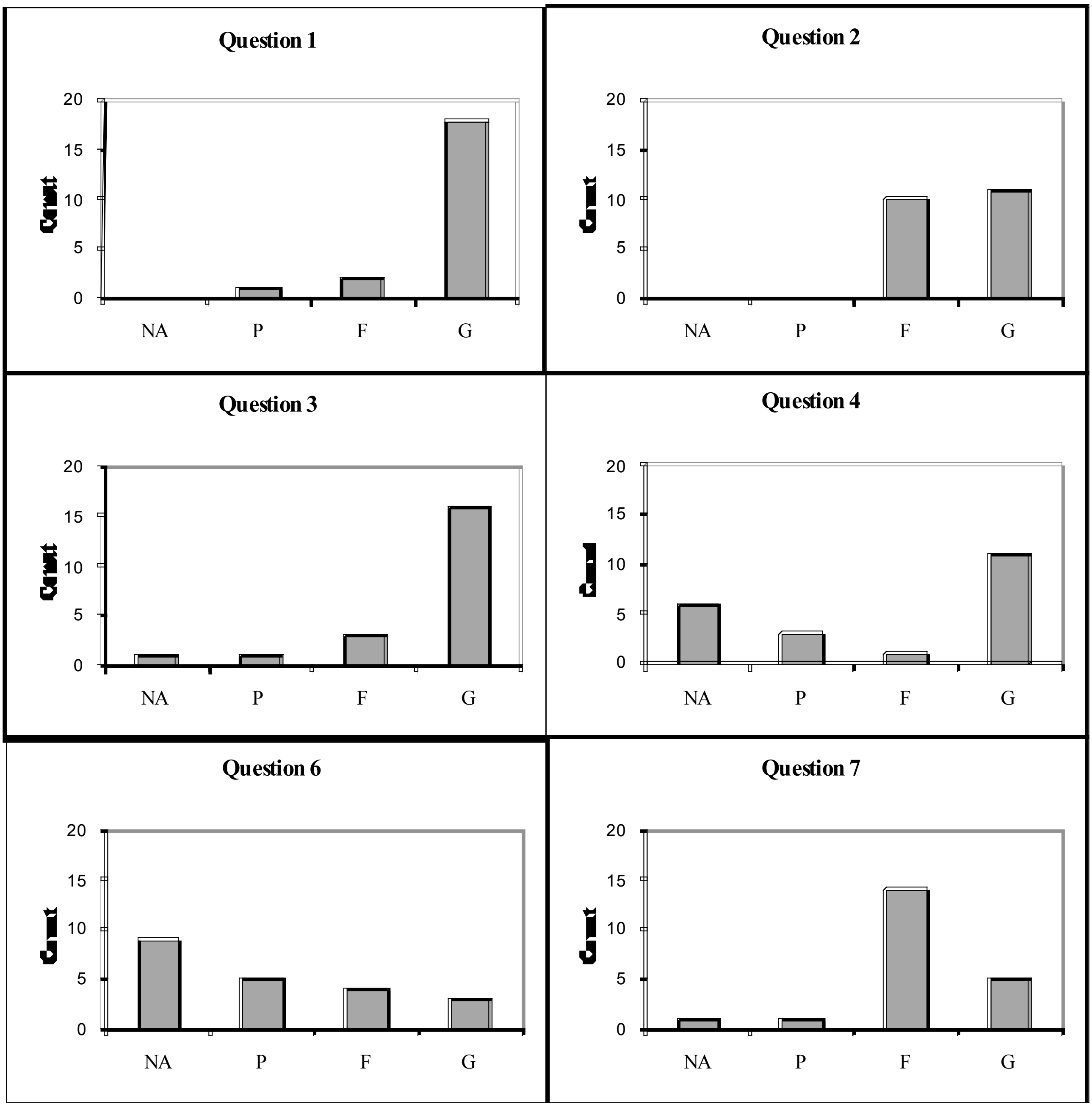

Figure 2. Assessment of the *Questions* of the activity with Non-Answered (NA), Poor (P), Fair (F) or Good (G).

FINAL REMARKS

As Fauvel and van Maanen (2000) point out, one should not underestimate the difficult task of the teacher to achieve a proper transmission of historical knowledge into a productive classroom activity for the learner. Given our lack of expertise in the field, in this first experience we were not able to foresee all the possible obstacles in the understanding process. Now we are aware of some difficulties inherent in the material (for instance, in *Questions 5* and *6*). First of all, the mathematical language and form (notation, computational methods, etc) turned out to be rather confusing right from the beginning. In addition, the syllabus and a sense of lack of time made us cram the activity into a two-hour class. Likewise, we had a slight doubt about how useful the topic was for our students. Why not give the opportunity to appreciate the topic in itself, stressing the aesthetics, the intellectual curiosity, or the recreational purposes involved? Finally, we borrowed and adapted part of Katz and Michalowicz's historical modules on Statistics, but in keeping with our syllabus, more didactic resource material on this topic should be elaborated for future use.

On the whole, however challenging, the experience proved to be rewarding in the end. Not only did the activity supply a collection of non-artificially designed problems, but it also helped to develop further skills, such as reading, writing and documenting. Above all, it was a means to show the original motivation of the normal curve and hence, to render it more understandable. This experience has shown that probability cannot be regarded as a collection of "polished" products within a deductive structured system, but rather as a system with a peculiar life (expectations, false expectations and false starts), as Guzmán (1993) put it, determined and influenced by external factors and connected with mathematical and non-mathematical domains.

REFERENCES

Blanco, M. & Ginovart, M. (2008). La probabilidad y la utilización de la plataforma virtual Moodle en las enseñanzas técnicas dentro del marco del Espacio Europeo de Educación Superior. Libro de actas del XVI Congreso Universitario de Innovación Educativa en las Enseñanzas Técnicas. Universidad de Cádiz.

Bloom, B.S. (ed.) (1956). Taxonomy of Education Objectives: Handbook I: Cognitive Domain. New York: David McKay Company [Major Categories in the Taxonomy of Educational Objectives]

Ellington, R. (1998). The importance of incorporating the history of mathematics into the Standards 2000 draft and the overall mathematics curriculum. EDCI 650 Reacts: History of Mathematics. University of Maryland.

Fauvel, J. & van Maanen, J. (eds.) (2000). History in mathematics education: the ICMI study. Dordrecht: Kluwer.

Garfield, J. (1994). Beyond Testing and Grading: Using Assessment to Improve Student Learning. Journal of Statistics Education [Online], 2(1). [Journal of Statistics Education, V2N1: Garfield]

de Guzmán, M. (1993). Origin and Evolution of Mathematical Theories: Implications for Mathematical Education. Newsletter of the International Study Group on the History and Pedagogy of Mathematics, 8 (March), 2-3.

Katz, V.J. & Michalowicz, K.D. (eds.) (2005) Historical Modules for the Teaching and Learning of Mathematics. Washington: The Mathematical Association of America.

Pengelley, D.J. (2002). A graduate course on the role of history in teaching mathematics. In Otto Becken (eds.). Study the Masters: the Abel-Fauvel conference. Gothenburg: National Center for Mathematics Education, University of Gothenburg [http://www.math.nmsu.edu/~davidp/gradcourserolehist.pdf]

Radke-Sharpe, N. (1991).Writing As a Component of Statistics Education. The American Statistician, 45, 292-293.

Swetz, F.J., Fauvel, J., Bekken, O., Johansson, B. & Katz, V. (eds.) (1995). Learn from the Masters. Washington: The Mathematical Association of America.

Tzanakis, C. & Arcavi, A. (2000). Integrating history of mathematics in the classroom: an analytic survey. In Fauvel & van Maanen (2000), 201-240.