Introducing a Study and Research Path in a course of Statistics for Engineering

Maria-Josep Freixanet, Montserrat Alsina, Marianna Bosch

INCORPORATING A STUDY AND RESEARCH PATH INTO A STATISTICS COURSE FOR ENGINEERING STUDENTS

Maria-Josep Freixanet[1], Montserrat Alsina[1], and Marianna Bosch[2]

[1]Universitat Politècnica de Catalunya, Mathematics Departament, Spain, maria.josep.freixanet@upc.edu; [2]Universitat de Barcelona, Spain

This paper describes a teaching proposal in engineering education based on the introduction of a study and research path (SRP) in the subject of Statistics of a Bachelor's Degree in ICT Systems Engineering. The conditions in which this SRP was designed and the first steps of its implementation are explained. Since an SRP starts with a generating question, special attention was paid to how the students worked on its formulation and the corresponding construction of the derived questions. With respect to the statistical knowledge mobilised, the problem of finding appropriate data in a format that can be applied appears as a key step to determine the study suitability of the generating question.

Keywords: study and research paths, statistics, anthropological theory of the didactic.

INTRODUCTION

Research on the implementation of *study and research paths* (SRPs) in university education is one of the developing areas of research in the Anthropological Theory of the Didactic (Barquero et al., 2021; Chevallard, 2015). SRPs appear as a specific type of enquiry-based teaching proposal that aims to make the prevailing pedagogical paradigm evolve from a knowledge-based study method towards a question-centered form of study. Their specificities compared to other forms of enquiry-based teaching formats are analysed in (Markulin et al., 2021). Various modalities of integrating SRPs into current university teaching have been explored (Bosch et al., 2020), together with the specific instructional devices implemented and the institutional constraints found. These constraints are mostly related to the prevailing pedagogical paradigm and the critical changes in the traditional didactic contract required by the new instructional proposal.

We present an experience related to the implementation of an SRP in a first-year course of Statistics for engineering students. Two previous studies present similar teaching proposals: three editions of an SRP in a Statistics course for second-year Business Administration students (Markulin et al., 2022), and an SRP in a Statistics course for Chemistry Engineering students (Quéré, 2022). Even if all the proposals correspond to the same theoretical model materialised in an SRP, numerous differences in the educational institutions, type of studies, course syllabi, students, group size, duration, form of implementation, etc. are observed. What they have in common is that the SRPs start from the consideration of an open question proposed by the teachers or by an external organisation – for instance a client or a firm – about

the professional environment of the studies: a marketing study in the first case, quality control of pharmaceutical products in the second.

This paper presents a research study in process that is also based on the implementation of an SRP in a university Statistics course including different characteristics that will be explained later. The productivity of approaching a real question is taken from (Markulin et al., 2022) and (Quéré, 2022). In this respect, SRPs (as other project-based or enquiry-oriented instructional proposals) lead students to encounter many aspects of data analysis that are usually absent from traditional classroom statistics practice, like data gathering, cleaning, and sorting. The current research relies on some findings of previous studies about SRPs for engineering students like the use of questions-answers maps as a tool to describe, share and manage the steps and components of the study process followed (Bartolomé et al., 2019; Florensa et al., 2018), and the importance of the situation and external contract in which the initial question is formulated (Barquero et al., 2021).

The novelty of this SRP is the delimitation of the open question that generates the SRP. Instead of presenting the students with a situation where an open question is raised (by the teacher or by an external organisation), they are provided with a topic and are asked to formulate questions they find interesting related to it. Our research question is the following: *is it feasible to start an SRP with an open topic and let the students decide upon the generating questions by themselves? What are the consequences of this choice?* Given the fact that the project takes place in a statistics course, it goes without saying that the questions raised will be addressed using data available or produced.

This paper describes the experience of the first step of the SRP in the collective construction of the generating question. We first present the analytical tools used, then the educational context of the experience. After that, we describe the activities implemented in class and the results obtained. The last section puts forward some learnings that can help better understand the potential of SRPs and the different forms they can take, together with the limitations of the options chosen.

THEORETICAL TOOLS FOR THE ANALYSIS

We will use the Herbartian schema proposed by Chevallard (2011) to identify some key elements of an SRP. Its reduced form is $S(X; Y; Q) \hookrightarrow A^{\heartsuit}$ and indicates a didactic system S where a group of students X, with the help of a group of teachers Y, addresses a question Q to provide their own answer $A^{\heartsuit}$. In this case, no generating question was provided, but a topic T that would lead to some questions Q_i requiring their corresponding answers $A_i^{\heartsuit}$, being the new schema $S(X; Y; T) \rightarrow < S(X; Y; Q_n) \hookrightarrow A_n^{\heartsuit} >$. The developed form of the schema $[S(X; Y; Q) \rightharpoonup M] \hookrightarrow A^{\heartsuit}$ includes a *milieu* M with all the resources used by $S(X, Y, Q)$ during the enquiry: questions Q_i derived from Q, external answers or works $A_j^{\diamondsuit}$ elaborated by others that seem useful to address Q, empirical data D_k and other pieces of knowledge, virtual and material objects O_m:

$$[S(X; Y; Q) \rightharpoonup \{Q_i, A_j^{\diamond}, D_k, O_m\}] \hookrightarrow A^{\blacktriangledown}.$$

In the case of this SRP, the schema could be described as follows:

$$[S(X; Y; T) \rightarrow < S(X; Y; Q_n) \rightharpoonup \{Q_{ni}, A_{nj}^{\diamond}, D_{nk}, O_{nm}\}] \hookrightarrow A_n^{\blacktriangledown} >.$$

Q_{ni} being the questions derived from Q_n, $A_{nj}^{\diamond}$ the answers or works that seem useful to address Q_n, together with empirical data D_{nk} and other pieces of knowledge, virtual and material objects O_{nm}. The implementation of SRPs produces important changes in the responsibilities assigned to the teacher and the students in the management of the different steps of the enquiry. These changes can be approached in terms of the evolution of the traditional *didactic contract* (Brousseau, 1997), and the tranfer of responsibilities from the teacher to the students.

DESCRIPTION OF THE INSTITUTIONAL SETTING

The *Universitat Politècnica de Catalunya* (UPC), like other European universities, has been going through a methodological transformation in education for some years, especially since the adoption of a competence-based curriculum within the European Higher Education Area. Educational innovation and student-centred methodologies have been developed in different areas, in particular since the implementation of the Research and Innovation in Learning Methodologies (RIMA) project by the UPC Barcelona Tech Institute of Education Sciences, and the foundation of the Journal of Technology and Science Education in 2011. Another significant boost was in 2017 with the creation of the Barcelona Science and Engineering Education (BCN-SEER), renamed to EduSTEAM, a research group in education mainly composed of senior researchers who, in addition to working in their areas of knowledge, had also been working on engineering education issues for some years (https://bcn-seer.upc.edu/en). After that, in 2020 the university was granted permission to offer a PhD programme in Education in Engineering, Science and Technology.

The SRP we are considering is part of the subject of Statistics offered in the second semester of the Bachelor's degree in ICT Systems Engineering in the Manresa School of Engineering – EPSEM, a campus of the UPC located one hour away from Barcelona. Although it is the first explicit implementation of an SRP at the UPC, it is some kind of evolution of previous experiences carried out by one of the authors in previous years (Alsina, 2022b). We can hence expect to find good conditions for its implementation.

Important institutional constraints also exist. UPC has engineering schools in Barcelona and in several nearby towns. However, they have a common Mathematics Department for most of them. The Manresa School of Engineering offers different Bachelor’s degrees: The degrees in Industrial Electronics and Automatic Control Engineering, in Mechanical Engineering, and in Chemical Engineering share all the first-year subjects, and the students are distributed into three groups that work in a coordinated way for each subject. The bachelor’s degree in ICT Systems Engineering has a different curriculum, but Statistics is offered in the second semester of the first

year too, which means it shares the syllabus, the assessment criteria and a final exam with the rest of the degree programmes.

At the syllabus level, the subject is organised into five main blocks, each one with a list of exercises and problems to be solved: descriptive statistics, probability and random variables, probability distributions, fundamentals of statistical inference and quality control, and main components. Assessment of the subject is based on two individual written exams (70%) and three individual activities to be done with the software used in class (30%). If necessary, the students will have to take a final exam, which is common for all the degrees.

Before this experience, other project-based activities had been carried out in the subject of Statistics by one of the authors, who is acting as an observer and researcher in this experience. Due to the organisation of the subject, those activities were implemented in the degree of ICT Systems Engineering, taking advantage of the software skills of the students. Following the same criteria, the SRP is also being developed in the same degree by the first author of the paper (who will be referred to as the teacher) with the approval of the Mathematics Department at EPSEM. The implementation coincided with the change of software used in the subject, from Minitab to R and R-Commander.

To decide on the assessment of the subject incorporating the SRP, we encountered the constraint that the final test (in case of not passing the continuous assessment) had to be the same for all the degrees, and we were asked to keep individual activities. Finally, the assessment was agreed upon as follows:

- SRP (20%): final report (10%) and a poster with including an oral presentation (10%).
- Two individual written tests (60%)
- Two individual activities to be done with the software used in class (20%)

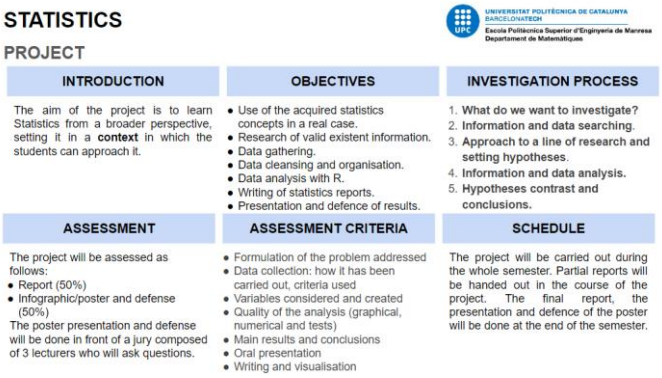
The SRP topic chosen was “water as an essential natural resource”, related to Education for Sustainability and the 2030 Agenda and its Sustainable Development goals (SDGs). This topic is linked to the UPC-led science dissemination project AquaeSTEAM, which seeks to propose scientific and technological questions, in which water is the common denominator. It is considered a challenge for students to explore, develop or create solutions, and find answers. On the website, educators and collaborators are provided with approaches, derived questions, resources, data and tools to work in the classroom (https://aquaesteam.upc.edu/).

To develop the SRP, the students were asked to work in teams of three or four members. The main operating tool used is the UPC virtual campus Atenea (a type of Moodle platform), in which the rest of the tools used in the project (padlet, Google Drive, Forms and Docs) are available. In this course, the lessons are structured into two weekly sessions of two hours each: one session with the entire group of students, and the other session with half the group, which is taught twice. The SRP is mainly carried out during the small group sessions, except when a pooling session is needed.

THE PROCESS OF FORMULATING A GENERATING QUESTION

In the very first session, for the students to gain insight into the whole subject, several activities were presented by the teacher. As an introductory activity, every student received a dice, and had to decide whether it was a trick dice or not. This activity follows a structure that allows the students to understand the subject, and have an overall perspective of what they are going to learn (Alsina, 2022a). It helps students to adopt a participatory role in the subject and work collaboratively, from the very first session, on an experimental activity directly related to statistics. It also allows them to gain self-confidence and use suitable vocabulary.

The SRP was then presented to the students: its objectives, the investigation process, the assessment criteria, and the schedule, as well as how it was going to be carried out: in teams of 3-4 students, and Atenea as the main tool used. See Figure 1 below for an image of the presentation of the course.


STATISTICS

PROJECT

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
Escola Politècnica Superior d'Enginyeria de Manresa
Departament de Matemàtiques

INTRODUCTION	OBJECTIVES	INVESTIGATION PROCESS
The aim of the project is to learn Statistics from a broader perspective, setting it in a **context** in which the students can approach it.	• Use of the acquired statistics concepts in a real case. • Research of valid existent information. • Data gathering. • Data cleansing and organisation. • Data analysis with R. • Writing of statistics reports. • Presentation and defence of results.	1. **What do we want to investigate?** 2. **Information and data searching.** 3. **Approach to a line of research and setting hypotheses.** 4. **Information and data analysis.** 5. **Hypotheses contrast and conclusions.**
ASSESSMENT	**ASSESSMENT CRITERIA**	**SCHEDULE**
The project will be assessed as follows: • Report (50%) • Infographic/poster and defense (50%) The poster presentation and defense will be done in front of a jury composed of 3 lecturers who will ask questions.	• Formulation of the problem addressed • Data collection: how it has been carried out, criteria used • Variables considered and created • Quality of the analysis (graphical, numerical and tests) • Main results and conclusions • Oral presentation • Writing and visualisation	The project will be carried out during the whole semester. Partial reports will be handed out in the course of the project. The final report, the presentation and defence of the poster will be done at the end of the semester.

Figure 1. Project presentation (our translation)

Step 1: formulating questions related to water

Inspired by water and its relationship with sustainability, climate change, or energy, the students were asked to propose and discuss some questions they were interested in. It allowed knowing the interest of the students, to make their priorities emerge, and get them used to the activity of questioning the reality and formulating questions. The students formed teams and started doing some research, gathering information and data related to water. They were asked to share their work with the rest of the group through a padlet (https://padlet.com/mariajosepfreixanet/hhlc43z62pqd0xo3) (see Figure 2). They started thinking of what caught their attention, what they wanted to know. This had to be formulated as a question, possibly the generating question of their project.

In the next session, with all the students, each team presented its research and its project proposal. Some of the project proposals were the following:

- What effects does society have on water?
- Has Covid-19 contributed to an increased water consumption of the population?
- What is the variation of water in the rivers of Catalonia every year?

- How to maintain a clean swimming pool and avoid fungi?
- Has water consumption increased in the last years?

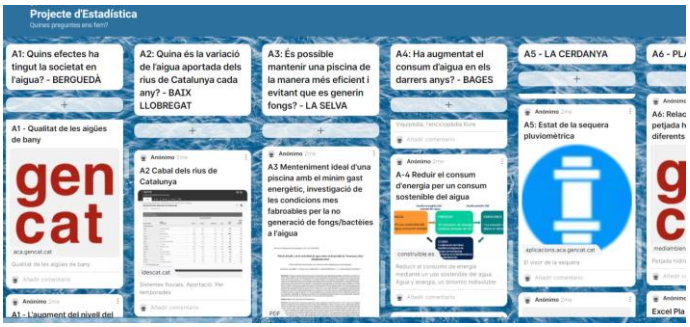


Figure 2. Padlet with the gathered information, data and project proposals

Step 2: linking questions with data

As a second step in the delimitation of the generating question, the teacher proposed the following task:

- *Write a short description of the aim of the study, the variables that are going to be analysed and why. What are these variables like?*
- *Include the link to a questions-and-answers map, which will be completed throughout the project.*
- *Enclose the data files you are going to use in your project.*

The teacher wrote a feedback report for each team and identified some difficulties most of the teams had when formulating the project proposals. In general, the questions were interesting, but some weaknesses appeared in the data provided:

a) Data format: wrong or difficult to analyse (five teams of 12 members)
b) Data content: not related or not answering the question (nine teams of 12 members)

With respect to (a), our interpretation is that the students lacked experience in the support software, and could not foresee the kind of data structure that was appropriate for the their project. As to (b), there may be two possible causes: difficulties in finding data, or in identifying statistical variables and their meaning.

A few examples of questions-answers maps (Q-A map) constructed by the students are included in Figure 3. It has to be noted that no Q-A map was provided by the teacher a priori, since the generating questions had to be given by the students. Nevertheless, the teacher explained what it was, the purpose of it and gave some examples. The teams only had to deal with the first level of the derived questions, without answers. Questions about the waste of water in Catalonia, or how the lockdown affected water consumption in Barcelona were topics of interest, but it is not easy to find data that can give accurate information about those issues.

Figure 3. Two examples of questions-and-answer maps

Step 3: analysing the project proposals

Instead of writing a feedback comment to each team, the teacher decided to explain some descriptive analysis concepts as well as the first steps of how to use R for the students to be aware of the difficulties related to their data, and to have a critical perspective. It was also to help them identify what type of data they could work with. This took three more sessions. The assessment of this first task was postponed.

Univariate and bivariate descriptive statistics concepts, as well as written examples and exercises in R were explained in these three sessions. A questionnaire about the students' previous contact with statistics was used to easily gather data and introduce the basic tools for its description and analysis (Figure 4).

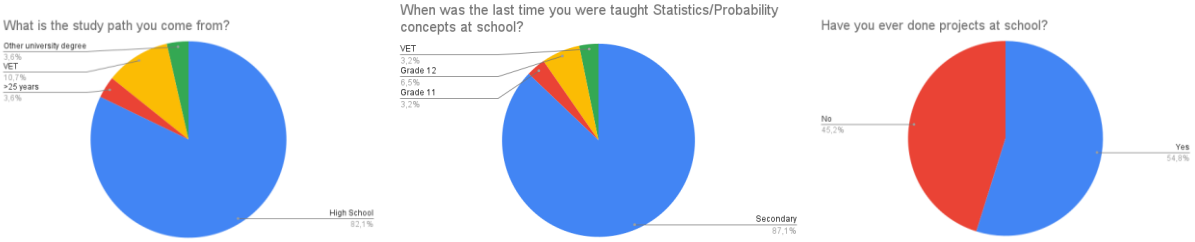


Figure 4. Results of the questionnaire on previous contact with Statistics

Once the descriptive statistics concepts and tools were introduced, each team of students had to analyse their own project proposal and the proposal of two more teams. A guide of questions was given to homogenise the analysis and to let the students think about the difficulties encountered by themselves. The questions to the students were asked are the following:

a) Data format: Can the provided data easily be analysed? YES/NO,
   Identified problem: The data are difficult to analyse, or have the wrong format
   - Can a descriptive study be implemented?
   - Can the data easily be uploaded to R?
   - Is it possible to find relationships between two variables? Which ones?
   - Is it possible to compare these variables with other data?

b) Data content: Can the provided data answer the main question? YES/NO
   Identified problem: The data do not answer the question

The students had to post this analysis as an answer to the project in the forum. Here are some of the results of the analysis of this activity, categorised by problems previously encountered by the teacher:

a) Data format: the data are difficult to analyse, or have the wrong format
Can the data easily be uploaded to R? • A5 by A3: "No Excel document is provided, so it would be annoying to gather all the information manually." • B3 by B3: "No, there is a lack of tables of values." • B3 by B1: "No, the data are not presented in tables." Can a descriptive study be implemented? • B3 by B2: "No, there is an important lack of data." • B4 by B3: "Information should previously be filtered. No tables of data are provided."
b) Data content: the data do not answer the question
• B3 by B2: "In the water footprint webpage, you have access to the total national consumption and the national consumption per capita. You can also find information on natality, general health, or income per capita and compare it to water consumption (https://waterfootprint.org/en/resources/interactive-tools/national-water-footprint-explorer/)" • A5 by A4: "the provided data indicate the water volume nowadays, a year ago, 5 and 10 years ago. It will not be possible to focus on the past 3 years". • A4 by A4: "we should search for new data about the water volume in reservoirs for more time periods. We could also consider a new variable about population increase, so we can analyse if there has been an increase of water consumption per inhabitant". • A3 by A1: "some more variables should be included: the presence of fungi, the quantity of fungi, the water volume, or the dimensions of the swimming pool (it may have a direct impact on how easily fungi grow) and if the swimming pool is private or public (public swimming pools tend to be bigger, water is more agitated and more people swim in them)".

At the end of this task, the students were asked to fill in a questionnaire to know their opinion of the analysis, and their perception of their learning process. When asked about their learning, the students mentioned the limitations of their initial proposals ("Unnecessary data were provided.", "The main question was too general.", "The data provided corresponded to many different years."), new aspects about data management and analysis ("Different approaches and how the others have used the data", "To carry out a good descriptive analysis and to select proper data", "To check if the revision and our data were correct"), and some improvement and ideas to move on with their project ("Possible mistakes we hadn't taken into account previously", "To find relationships between variables and inspiration for our project", "Aspects the other groups and our group can improve"). In total, seven sessions were used to implement steps 1, 2 and 3: two sessions for step 1, one session for step 2 and four sessions for step 3.

DISCUSSION AND CONCLUSION

The experience here presented focuses on the first step of the enquiry described by the Herbartian schema: $[S(X; Y; T) \rightarrow < S(X; Y; Q_n) \rightharpoonup \{Q_{ni}, A_{nj}^{\diamond}, D_{nk}, O_{nm}\}] \hookrightarrow A_n^{\blacktriangledown} >$. We observed how the topic of water (T) proposed to the students led to the

formulation of different generating questions Q_n and to the first steps of their study based on the identification of derived questions and the search of available answers $A_{nj}^{\Diamond}$ and data D_{nk}. The task proposed by the teacher in step 2 (linking questions with data) aimed at exploring the elements that could easily be integrated into the milieu during this phase of the enquiry. The teacher approached the students' difficulties by introducing new statistical tools and notions (O_{nm}) related to the concept of variable and the requirements of the statistical software used (R Commander). This work led the enquiry community to reject some of the proposed generating questions and to agree upon a few that appear to be suitable for the study – under the given conditions.

As far as the management of the SRP is concerned, the experience breaks some of the implicit clauses of the didactic contract, while maintaining others. Let us start with the first research question about the problem of choosing a productive generating question, which is an important issue related to SRPs (Markulin et al., 2021). By letting the students choose the SRP generating question Q_n instead of directly proposing it, a new responsibility is passed on to the students and, in a way, assumed by the entire study community composed of the students and the teacher. What is also shared with the students is the process of analysing the productivity of the questions raised and the viability of their study (their "studiability") concerning an important topic: the kind and quality of available data. This is an aspect of statistics that remains even more in the shadows than data cleaning and management, at least in educational contexts, and it certainly deserves more attention. It can be concluded that this way of starting the SRP has given the students the opportunity to learn the data suitable to be analysed statistically and the characteristics those data should have. However, this strategy took up more sessions than expected and represented an effort for the teacher to redirect the SRP. It is interesting to see how the use of questions-answers maps appears as a good strategy in this context.

Several aspects of the study process remained under the sole responsibility of the teacher. This is more than likely due to the prevalence of the traditional didactic contract. These aspects include the organisation of the tasks, the introduction of new concepts and tools, the pooling of results, and the planning of the sessions.

An important consequence shown in this paper is how much the movement of responsibilities from the teacher to the students results in the motivating and learning of a new type of statistical knowledge that would rarely appear otherwise. The conditions of "studiability" of the initial questions generating an SRP appear as a critical issue for the implementation of enquiry-based teaching proposals. Their approach does not seem to be only a question of the SRP design – therefore under the responsibility of the teacher – but of the SRP management – by both the teacher and the students.

ACKNOWLEDGEMENTS

The research leading to these results received funding from the Spanish R&D Project PID2021-126717NB-C31 (MCIU/AEI/FEDER, UE).

REFERENCES

Barquero, B., Bosch, M., Florensa, I., Ruiz-Munzon, N. (2021) Study and research paths in the frontier between paradigms. *International Journal of Mathematical Education in Science and Technology*, *53*(5), 1213-1229.

Bartolomé, E., Florensa, I., Bosch, M., & Gascón, J. (2019). A 'study and research path' enriching the learning of mechanical engineering. *European Journal of Engineering Education, 44*(3), 330-346.

Bosch, M., Barquero, B., Florensa, I., Ruiz-Munzon, N. (2020). How to integrate study and research paths into university courses? Teaching formats and ecologies. *INDRUM 2020*, Université de Carthage, Université de Montpellier. ⟨hal-03113981⟩

Florensa, I., Bosch, M., & Gascón, J. (2018). Enriching engineering education with didactics of mathematics: Study and research paths in engineering education. *IEEE Global Engineering Education Conference* (EDUCON) (pp. 751-759).

Alsina, M. (2022a) An inquiry to generate in a natural way Statistics syllabus: Is it just a matter of luck? (submitted).

Alsina, M. (2022b). Questions used as a compass: an experience in engineering degrees (submitted).

Bosch, M. (2018). Study and research paths: a model for inquiry. In B. Sirakov, P. N. De Souza, M. Viana (Eds.), *Proceedings of the International Congress of Mathematicians*, *ICM 2018,* vol. 3 (pp. 4001-4022). World Scientific Publishing.

Brousseau, G. (1997). *Theory of didactical situations in mathematics: didactique des mathématiques (1970-1990).* Kluwer Academic Publishers

Chevallard, Y. (2015). Teaching Mathematics in tomorrow's society: a case for an oncoming counter paradigm. In S. J. Cho (Ed.) *Proceedings of the 12th International Congress on Mathematical Education* (pp. 173-187). Springer.

Markulin, K., Bosch, M., & Florensa, I. (2021). Project-based learning in statistics: A critical analysis. *Caminhos da Educação Matemática em Revista*, *11*(1), 200-222.

Markulin, K., Bosch, M., & Florensa, I. (2022). Study and research paths in statistics: an ecological analysis. *Proceedings of CERME12* (in press).

Quéré, P. V. (2022). Bridging the mathematics gap between the engineering classroom and the workplace. *International Journal of Mathematical Education in Science and Technology, 53*(5), 1190-1212.