

A pilot study and research path in statistics at the university level

Janielly Verbisck, Laura Fernández-Ruano, Nathan Lombard

► To cite this version:

Janielly Verbisck, Laura Fernández-Ruano, Nathan Lombard. A pilot study and research path in statistics at the university level. Fifth Conference of the International Network for Didactic Research in University Mathematics, Jun 2024, Barcelona, Spain. pp.513-522. hal-04912364

HAL Id: hal-04912364 https://hal.science/hal-04912364v1

Submitted on 26 Jan 2025 $\,$

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

A pilot study and research path in statistics at the university level

Janielly Verbisck¹, Laura Fernández-Ruano², and Nathan Lombard^{1,2} ¹University of Barcelona, <u>janielly.verbisck@ub.edu</u>, Spain, ²IQS, Ramon Llull University, Spain

This paper reports on a university teaching experience in statistics based on a study and research path (SRP), an inquiry-based instructional proposal elaborated within the Anthropological Theory of the Didactic (ATD). The SRP corresponds to a pilot case for an ongoing dissemination project aiming at implementing instructional proposals close to the paradigm of questioning the world in the university context. Its description illustrates the ATD characterisation of SRPs in terms of the Herbartian schema and its dialectics. It particularly focuses on the didactic infrastructure created by the lecturer and the elements of the instructional strategy that could be disseminated to other lecturers in other university settings.

Keywords: study and research paths, Herbartian schema, dialectics, inquiry-based teaching, statistics education.

INTRODUCTION

Nowadays, university education receives a lot of pressure to shift from a teachercentred pedagogy based on the frontal study of pieces of knowledge towards more student-centred and inquiry-based approaches to teaching and learning. However, pedagogies of the latter type require the implementation of conditions that are sometimes difficult to reach (Markulin et al., 2021). As a result, a lot of inquiry-based teaching proposals do not overcome the experimental stage. When implemented under different institutional conditions, they tend to vanish or take on forms more consistent with the old "frontal" pedagogy (Markulin et al., 2022).

This issue may be characterised in the framework of the Anthropological Theory of the Didactic (ATD) (Chevallard & Bosch, 2020), which puts it in terms of a shift in paradigm from the *paradigm of visiting works* to the emerging *paradigm of question-ing the world* (Chevallard, 2015). Theorizing on the latter paradigm has allowed ATD-researchers to develop the pedagogy of *study and research paths* (SRPs) (Bosch, 2018; Chevallard, 2015). The implementation and analysis of SRPs in different institutional contexts brought important local results about the conditions needed for SRPs to be integrated into educational institutions and the constraints hindering their development and dissemination (Barquero et al., 2022). For the conditions needed, three levels may be distinguished (Markulin et al., 2022): the epistemological related to the structure and organisation of the content that is to be taught and learnt, the didactic one related to the way this content is managed during the teaching and learning processes, and the pedagogical related to the strategies and devices that are general to different content, domains, and disciplines. In this paper, we focus on the didactic level.

LABINQUIRY (Lombard et al., this issue) is a project to transfer research results about SRPs to the secondary school and university levels. It aims at providing the necessary didactic infrastructure for the design and implementation of SRPs. It consists of the development of a website as well as a set of online didactic modules adaptable to learning platforms (such as Moodle and Google Classroom), together with the creation of a community of lecturers all implementing SRPs. LABINQUIRY will feature prototypical SRPs with their potential generating question, a priori analysis, collections of structured resources and analyses from previous experiences, links to external resources and potential experts, etc. In this context, the LABINQUIRY research team selected two SRPs that have been developed and tested throughout the years, as pilot studies for the transfer procedure. The first one is about combinatorics implemented many times at secondary schools (Vásquez et al., 2021) and the second one is an SRP in statistics for bachelor's degrees in chemistry, chemical engineering, and industrial technologies engineering (Fernández-Ruano et al., 2024).

In the present paper, we report on the university SRP, which is taught by the second author of this paper, and whom we will call P_1 in the following. The SRP taught by P_1 was implemented over the last four years adopting different forms. Each implementation is analysed and used to improve the next one. Therefore, the last version of the SRP is considered to be the greatest adapted to its institutional setting. In the following, we will describe how we observed it to give it the role of a pilot for the LABINQUIRY project. More precisely: *What didactic infrastructure is created for the SRP and how is it used or activated by the lecturer and the students in class?*

THEORETICAL FRAMEWORK

SRPs are long teaching and learning processes that start with the consideration of an open-generating question students address under the guidance of the teachers. The generating question is expected to be open enough to require the search for information from different sources and the study of this information to elaborate and validate a final answer, obtained collectively by the students under the guidance of the teacher(s). An important aspect of SRPs is the fact that the question approached should always remain the main goal of the inquiry, instead of being addressed as a pretext to introduce new concepts, knowledge organisations or tools.

SRPs can be described through the so-called *Herbartian schema*: $S(X; Y; Q_0) \rightarrow A^{\bullet}$, where a group of students X, helped by a group of teachers Y, form a didactic system S to address an initial question Q_0 and provide a final answer A^{\bullet} . In the process from Q_0 to the collective elaboration of A^{\bullet} , the *didactic system* $S(X; Y; Q_0)$ displays Q_0 into derived questions Q_i , searches already available "labelled" answers A_j^{\diamond} , elaborates and adapts them to Q_i , finds new questions during the process which, in turn, call for new answers, and so on. Bosch (2018) pointed out the importance of the questions and answers (Q-A) dialectic to ensure the dynamics of SRPs. The Q-A dialectic provides visible proof of the progress of the inquiry and contributes to the overall process management. To elaborate A^{\bullet} , the didactic system creates a didactic

milieu M: $[S(X; Y; Q) \rightarrow M] \rightarrow A^{\bullet}$. This *milieu* is composed of derived questions Q_i , "ready-made" answers A^{\diamond_j} that seem helpful to answer Q_i , any kind of works W_k (knowledge or material), and the sets of data D_m of all natures gathered during the inquiry. The extended *Herbartian schema* is symbolised as:

 $[S(X; Y; Q) \rightarrow \{Q_1, Q_2, ..., Q_i, A^{\diamond}_1, A^{\diamond}_2, ..., A^{\diamond}_j, W_1, W_2, ..., W_k, D_1, D_2, ..., D_m\}] \rightarrow A^{\bullet}$

The *media-milieu* (*Me-Mi*) *dialectics* becomes crucial during the whole SRP. To analyse this dialectic, we look at where external information, data and answers come from, and how their access is managed (*media*). We also ask how they are validated and transformed; and with what materials the final or intermediate answers are developed (*milieu*). Finally, an SRP is a collective inquiry process during which small groups X_i are formed and individual work is also carried out. X_i and Y_j should organize themselves to work together. Hence a necessary share of responsibilities must be constantly established, as to what questions should be studied, what strategy should the class (as a group) adopt, what answers are considered valid, and so forth.

So, when analysing an SRP, one may pay close attention to three principal aspects (Barquero & Bosch, 2015). First, the *chronogenesis* of how the teacher monitors the questions-answers dialectic; then the *mesogenesis* of how the teacher stages the media-milieu dialectic. Finally, the *topogenesis* corresponds to the position (*topos*) assumed by the teacher and the students and their sharing of responsibilities. In this paper, we focus on the analysis of these three aspects in the pilot SRP.

METHODOLOGY TO DESCRIBE THE PILOT SRP

This paper presents the analysis of the last version of the SRP (course 2023-2024), out of four implementations. The SRP is carried out in the first statistics subject of a bachelor's degree in chemical engineering, a subject taking place in the first semester of the second year. It corresponds to 6 ECTS credits and students have 39 hours of classes together. The first and last authors of this paper conducted non-interference observations of the SRP's sessions and in some of the lectures to observe the P_1 's management and the students' activity. Then, they interviewed P_1 to shed light on aspects of the preparation and the management of the SRP we could not grasp through our sole observations. We also asked P_1 about the changes made since her first edition of the SRP (2020-2021) until the last version in progress.

The data gathered consisted of our written records of the non-interference observations in the SRP sessions, the transcribed interview with P_1 and the materials made available by P_1 on the SRPs previously implemented. With these data, we particularly focus on P_1 's strategy to manage the SRP's *chronogenesis, mesogenesis*, and *topogenesis*. So, we described these three aspects in terms of the Herbartian schema, the Q-A and the media-milieu dialectics. We also described the general organization of the SRP and the role of an external instance as a possible element in promoting changes in the didactic contract in the current SRP. These two moments were articulated with P_1 's *topos*, which has an essential position in both the running of the SRP and the development of the LABINQUIRY project.

These pieces of evidence will enable the dissemination procedure of the SRP didactic infrastructure as a "model device" that will be available in the LABINQUIRY project to lecturers who want to develop it in their practice.

THE GENESIS OF THE SRP AND ITS INTEGRATION INTO THE SUBJECT

 P_1 is a lecturer with a long experience and who in recent years dedicated her teaching to topics of statistics in chemical engineering courses in a private university in Barcelona. In 2020, she learnt about the SRP device through an ATD researcher and began to design and implement it in her statistics courses.



Figure 1: Generative questions of the previous editions of the SRP designed by P_1

In the first edition (2020-2021), P_1 proposed an SRP as an extra class work. Previously, a data set about the concentration of air contaminants in Catalonia was found by her, and she aimed her students to work on these data. She gave the students the data set and requested them, organized by groups, to raise questions about the air quality and try to give an answer after organizing, representing, and analysing the data (Fernández-Ruano et al., 2024). In the second edition (2021-2022), the students had to work with the same data set but this time the SRP was part of the statistics subject, and P_1 organised the sessions to introduce statistics knowledge and sessions dedicated to the development of the SRP. In the third edition (2022-2023), P_1 structured the SRP to run in parallel to the lectures. All the generating questions are presented in Figure 1. In all these editions, the common point was the questioning about the air quality and the accessibility of a data set regularly updated by the Catalan government about the information collected at several observation points in the territory. Evolutions over time appear because each edition of the SRP provided elements of the *a priori analysis* of the next one leading P_1 to make changes not only to the generating question but also to the didactic infrastructure and its management.

In the last edition in 2023-2024, P_1 implemented the SRP in the subject of statistics for a second-year bachelor's degree in chemical engineering. This subject focuses on the fundamental statistical principles of the field of chemical engineering. Its syllabus highlighted the development of abilities to: recognize, create, and resolve chemical engineering problems that call for the application of statistical approaches; disseminate knowledge, concepts, issues, and solutions to both specialized and general audiences; and solve computational issues and statistical analysis. The general organisation of the entire subject is presented in Table 1. It is important to highlight the relationship between the statistics lectures and the SRP and how this last one gives a raison d'être to the subject of statistics, in line with the proposal by Barquero et al. (2018) about linking transmission with inquiry at the university level.

Week	Brief description of the lectures	Brief description of the SRP
1	Presentation of the subject: What is statistics? Objectives of statistics. Uncertainty of the data and core idea.	P0. Presentation of Q_0 by the external instance. P1a. Viewing the dataset about air contaminants. Autonomous work: Organization of the groups. Derived questions raised by groups in the diary.
2	Exploratory data analysis. Construction and interpretation of statistics, tables, and charts.	P1b. Organizing the derived questions and discussion to select the relevant ones. Autonomous work: search for "ready-made" answers A^{\diamond}_{j} in different media; report in the diary.
3	Distributions: from sample to popu- lation. Definitions and models of frequent use.	P2. Revision of answers A^{\diamond}_{j} and new questions. Autonomous work: continue to search for infor- mation on the assigned questions, and report in the shared document and in the team diary.
4	Inference: introduction and parame- ter estimation. Inference techniques. Sample distri- butions. Interval estimation of the mean and variance. Sample size.	P2. General discussion of the selection of variables and first organization of data. Autonomous work: continue to search for information on the assigned questions, and report in the shared document and in the team diary.
5	Hypothesis testing for one and two samples. Parametric and non- parametric tests.	Bank holiday (no work in the classroom). Autonomous work: Submission of pre-report 1 by teams (introduction and first data representa- tion).
6	Hypothesis testing for one and two samples. Parametric and non- parametric tests.	P2. Review of pre-report 1: general discussion about agreements and pending issues <i>Autonomous work: Integrate the feedback pro-</i> <i>vided and fill in the team diary.</i>
7	Hypothesis testing for one and two samples. Parametric and non- parametric tests.	P3. Representation and description of the orga- nized data. Autonomous work: Propose ques- tions to the external instance. Submit 2nd pre- report.
8	Frequency tables. X^2 test for independence. Goodness of fit test.	Partial exam.
9	Analysis of Variance. Sources of variation. One-way ANOVA.	P3. Final data selection: general agreements. Autonomous work: teams make the assigned basic graphs and one of them unifies the graphs.
10	Analysis of Variance. Sources of variation. One-way ANOVA.	P4. Teams work on internal answers: graphs and data description, inferential analysis, conclusions. Autonomous work: Groups work on pre-report 3 and on part of the final report.
11	Analysis of Variance. Post-hoc analysis. Two-way ANOVA. Non-	P4. Preparing posters.

	parametric models. Introduction to Design of experiments.	
12	Non-parametric models. Introduction to Design of experiments. Regression models.	P4 & P5. Presentation of internal answers (posters).
13	Regression models. Correlation as a measure of linearity. Least-squares method. Diagnostic graphics. Ade- quacy of the regression. Interpola- tion.	Bank holiday.
14	Regression models. Least-squares method. Diagnostic. Adequacy of the regression. Interpolation.	P5. Presentation of the final answer to the exter- nal instance.
15	Finalization of the subject. Final exam.	

Table 1: General organisation of the lectures and SRP sessions by weeks

In the previous editions, P_1 recorded the progress of the SRP at each session, but without the need for a well-defined terminology to demarcate the different inquiry stages. In this latest edition, it appeared the need to co-create a didactic infrastructure that will later be made available to other teachers through LABINQUIRY. We then saw the need to distinguish between the different stages of the inquiry that can occur during the implementation of the SRP. These stages are distinguished in terms of phases (P0, P1a, P1b, P2, ..., P5) and have been highlighted in Table 1.

THE PILOT SRP AND ITS MANAGEMENT

Chronogenesis

The generating question of the SRP is Q_0 : Are the low emission zones (ZBE) correctly dimensioned and to what extent do their dimension affect their efficiency? During phases P0-P2, the students raised derived questions, grouped the questions (location, legislation, temporal, pollutant, and working with data), and discussed and selected the relevant ones for the study. They then selected the data to work during phases 2-4. The derived questions studied during the different phases were:

Phases 0-2

Q1.1: Which areas of Catalonia are ZBE, where exactly are they located?

Q1.2: In the pollutant emission zones, are emissions measured in 2D or 3D? How is the extent and geographical radius of pollution defined?

Q1.3: Has there really been a decrease in the level of pollutants in the established ZBE?

Q1.4: Should all settlements with more than 5000 inhabitants be taken into account?

Q1.5: What percentage of zones are ZBE compared to the total number of zones in Catalonia?

Q1.6: Which population to select? Which stations should be selected?

Q2: Since when do the ZBE function (are sanctions imposed)?

Q3: What is the optimal time period to evaluate the effectiveness of the ZBE and should several years of data be considered?

Q3.1: Which hours, days, months, and years to select? How to select the data to obtain this information?

Q4: Which are the most important parameters (NOx, SOx, microparticles, ...) which affect the delimitation of the ZBE?

Q4.1: Are polluting gases differentiated by the degree of pollution they produce or are they treated equally?

- Q4.2: Which pollutants are related to traffic?
- Q4.3: Does meteorology affect the mobilisation of pollution?
- Q4.4: Which pollutants to select?
- Q4.5: What are the differences between primary and secondary pollutants?

Phases 3-4

Q5: How do you start working with the data provided at the statistical level?

Q5.1: How do you measure effectiveness?

Q5.2: What criteria will we use to choose the data? How should the data in the table and its analysis be related to the initial question?

- Q5.3: How are the data taken from the available database?
- Q5.4: How often are data on air emissions collected?
- Q5.5: What variables are to be considered?
- Q5.6: What would be the next step to move forward with the work?

During almost every session, we could see P_1 arranging some time to discuss a map with the questions addressed during the SRP with the students. Especially, she repeatedly asked about the status of the proposed answers to each question: whether they should be taken for granted, be further investigated (rather trusting or rather disproving them) or be dismissed. Then, the decision was on the students' side (see below). An important aspect regarding the animation of the dialectic of questions and answers was also the availability of data, which is a critical point in statistics.

Mesogenesis

During all the phases of the SRP development, students were guided by P_1 to consult numerous types of media. At the same time, the data accessed provided a rich *milieu* for the evolution of the inquiry. The resources available mainly consisted of existing A_j^{\diamond} answers found on the Internet and incorporated by the groups into their *milieu*, D_m data on air pollutants made available by P_1 (which, in fact, is the database used in all previous editions of this SRP), a summary of the derived questions grouped and selected collectively, and knowledge tools W_k provided by P_1 throughout her classes in statistics. Once again, the relationship between the statistics lectures and the development of the SRP is highlighted when P_1 , in her lectures, provides students with the study of data in different contexts and indications of how to work with *Excel* for the organisation, representation and analysis of these data. Therefore, the new knowledge necessary to carry out the SRP is progressively introduced by her during the lectures.

In general, the validation of the answers is done through collective discussions about the answers elaborated by the groups. P_1 organises SRP sessions to work in groups

(for the elaboration of partial answers) so that collective discussion sessions may happen afterwards to make decisions on the next steps to follow. The following excerpt can illustrate this type of discussion:

 P_1 : The other day I asked x_1 "Are you not going to discuss anything about this? Because he had been saying for three or two months that "meteorology had to be taken into account"... He wrote an impressive report on why it had to be taken into account and when the day came to ask "shall we eliminate it?", he kept quiet and I said " x_1 , now you're not going to tell anything about this, you're not going to discuss it?", and he said "no".

All the groups' productions are presented in diaries and pre-reports. Many of the productions are shared in cloud storage files, a collective means created by P_1 so that everyone has access to the groups' productions. So, P_1 facilitates working sessions in a variety of places. This expansion goes with a clear increase in the possible milieus of study, some being directly animated by the lecturer, others not. This is how she describes the infrastructure used:

 P_1 : This year, the crucial difference is that we are combining work in groups in the tutorial room, with collective work in the lecture room, together with homework. So we use these three spaces. [...]

Topogenesis

An important aspect of the topogenesis, besides the new responsibilities students must assume (raising questions, searching for information, elaborating answers, etc.), is related to the use of students' intermediate answers. Every week, P_1 gathered the answers A_i^{\bullet} provided by the groups in their reports. These answers then became new pieces of work A_i^{\diamond} for the general inquiry. We see here how the questions and answers dialectic is entangled with media-milieu dialectic, as students are considered (first by P_1 but also by themselves) as legitimate media whose productions deserve to be discussed. This enforces their topos, because the students' results are not only intended to the lecturer but to the entire inquiry community.

Another important reinforcement of the topogenesis has been produced by the intervention of an external agent to propose the generating question. In this edition, it was a representative of the Environmental Department of the Government of Catalonia who was invited to present the generating question to the students as a request of the traffic section he heads. In the three previous editions, the generating question was presented by P_1 as a work that was part of the statistics subject. The students took it as an exercise and did not take the study so seriously. In this latest edition, a big change is noticeable because the final answer is no longer a simple exercise, but a response to be presented to a real request. P_1 made it clear that the importance of the external agent was in "the responsibility" transferred to the students. Of course, this also changed the final product of the inquiry A^{\bullet} : P_1 : Somehow we have to reach a consensus, we have to give an answer and this, for me, has opened up the world. You know they see that they must give an answer to a question that is not an exercise, that changes things.

In fact, this change did not only affect the nature of the expected answer but also the level of expectation. That is, such an intervention manifestly enforced the necessity to answer properly, meaning intelligibly and argumentatively.

CONCLUSION

This paper presents the methodology used to characterise an SRP that has been implemented during four consecutive years and acquired a certain maturity so as to be used as a pilot inquiry process for a dissemination project. The main elements used are the Herbartian schema and the dialectics of questions-answers (chronogenesis), media-milieu (mesogenesis) and individual-collective (topogenesis). The results obtained by the a priori and in vivo observation and analysis of the inquiry process point at a coherent and self-sustained instructional practice, rooted in a rather stable didactic infrastructure. Some relevant aspects to highlight affect the three dialectics. The strategy to manage the chronogenesis alternates sessions of teamwork and collective discussions while establishing a close connection between the content of the lectures and the demands produced by the SRP. What is visible in the evolution of the SRP during the three years of implementation is a radical change of the statistical content of the course, which is increasingly aligned with the needs generated by the SRP. In this respect, we can notice a rather strong transformative power of the SRP, which brings an interesting perspective on the ecology of SRPs and their potential dissemination. The didactic infrastructure created by the lecturer to support the mesogenesis includes an elaborated organisation of online resources to facilitate the students search for data and new pieces of knowledge and their sharing. The summaries of students' productions prepared by the lecturer nourish the inquiry dynamics and transform the intermediate answers in new pieces of information to be validated by the inquiry community. This strategy reinforces the topogenesis in the responsibility transferred to the students about the relevance of their productions and those of their classmates. Finally, the intervention of an external agent to present the inquiry generating question appears as a decisive factor for the students' assumption of responsibilities in the entire process and let the generating question assume a leading role during the inquiry. The fact that the lecturer is not an expert in didactics but has managed to contribute to the creation of innovative resources to sustain the SRP opens positive expectations for the ongoing dissemination project in other university settings.

ACKNOWLEDGEMENTS

Funded by Projects PDC2022-133812-C21, PID2021-126717NB-C31 and PID2021-126717NB-C33 (MCIN/AEI and FEDER A way to make Europe).

REFERENCES

- Barquero, B., & Bosch, M. (2015). Didactic engineering as a research methodology: from fundamental situations to study and research paths. In *Task Design In Mathematics Education. New ICMI Study Series* (pp. 249–272). Springer. https://doi.org/10.1007/978-3-319-09629-2 8
- Barquero, B., Bosch, M., Florensa, I., & Ruiz-Munzón, N. (2022). Study and research paths in the frontier between paradigms. *International Journal of Mathematical Education in Science and Technology*, 53(5), 1213–1229. https://doi.org/10.1080/0020739X.2021.1988166
- Barquero, B., Monreal, N., Ruiz-Munzón, N., & Serrano, L. (2018). Linking Transmission with Inquiry at University Level through Study and Research Paths: the Case of Forecasting Facebook User Growth. *International Journal of Research in Undergraduate Mathematics Education*, 4(1), 8–22. https://doi.org/10.1007/s40753-017-0067-0
- Bosch, M. (2018). Study an research paths: a model for inquiry. In B. Sirakov, P. de Souza, & M. Viana (Eds.), *International congress of Mathematicians* (Vol. 3, pp. 4001–4022). World Scientific Publishing Co. Pte. Ltd. https://doi.org/10.1142/9789813272880_0210
- Chevallard, Y. (2015). Teaching Mathematics in Tomorrow's Society: A Case for an Oncoming Counter Paradigm. In S. J. Cho (Ed.), *The Proceedings of the 12th International Congress on Mathematical Education* (pp. 173–187). Springer . https://doi.org/10.1007/978-3-319-12688-3_13
- Chevallard, Y., & Bosch, M. (2020). Anthropological Theory of the Didactic (ATD). In *Encyclopedia of Mathematics Education* (pp. 53–61). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0 100034
- Fernández-Ruano, L., Serrano, V., Báguena, J., Cuadros, J., Sáez, M. (2024). A practical approach to an SRP in university Statistics. In I. Florensa et al. (in press). *CITAD7 Extended Abstracts 2022*. Springer Birkäuser & CRM.
- Markulin, K., Bosch, M., & Florensa, I. (2021). Project-based learning in statistics: a critical analysis. *Caminhos Da Educação Matemática Em Revista*, 11(1), 200–220.
- Markulin, K., Bosch, M., Florensa, I., & Montañola, C. (2022). The evolution of a study and research path in Statistics. *EpiDEMES*, *1*. https://doi.org/10.46298/epidemes-7584
- Vásquez, S., Barquero, B., & Bosch, M. (2021). How Long Would It Take to Open a Padlock? A Study and Research Path with Grade 10 Students (pp. 105–115). https://doi.org/10.1007/978-3-030-76413-5_12