Research-based learning: a case study for engineering students

Julieta Noguez¹ · Luis Neri¹

Received: 16 January 2019 / Accepted: 2 May 2019 / Published online: 20 May 2019
© The Author(s) 2019

Abstract

An implementation of the research-based learning (RBL) model and methodologies for undergraduate Computational Engineering students at Tecnológico de Monterrey, Mexico City Campus was undertaken to highlight the importance of involving undergraduate students in professional research activities early in their career (curriculum). We present here the results of this study. Four phases of the project are described: (1) a diagnosis of initial RBL competencies of students; (2) the presentation of an introductory workshop on research methodology by the teacher; (3) the design of a monograph by the students; and (4) the preparation of a research project by the students, including a sketch of a research paper and a written and oral presentation to a jury of staff research professors. The objective of the project was the design and evaluation of algorithms of Data Science, Big Data and Deep Learning to perform a data processing analysis related to the problem of metabolic syndrome and its surrogates in Mexico. The results indicate that the initial research competencies of undergraduate students are insufficient, although about 50% of the students think the contrary. The students were able to work in teams during the semester, and most teams, guided by the instructor, were capable of developing appropriate monographs and high-quality sketches for their proposed research papers. In most cases, students showed great improvements in their RBL skills in doing extensive literature reviews and in performing appropriate statistical analyses of databases. This work supports the contention that it is particularly important to start developing RBL competencies early in undergraduate engineering careers in order to prepare undergraduate students for future graduate studies as well as for their professional lives.

Keywords Didactic techniques · Educational innovation · Higher education · Learning activities · Research based learning · Research competencies

1 Introduction

Research-based learning (RBL) aims to promote and develop student competencies related to research practice and to benefit students through activities linked to research [1]. This technique implies the application of learning and teaching strategies that link research with teaching.

Given that the reach and implementation of RBL for undergraduate students is very wide, it is necessary to define the learning objectives, the audience, the expected learning outcomes and the desired competencies to be developed in

order to determine the appropriate methodology and processes to be applied.

The current trend for engineering education invites to establish interactive and active activities for a deeper, more comprehensive and far-reaching learning by students. In this sense, various methodologies have been proposed for active learning based on the search for information (inquiry-based learning) such as problem-based learning (PBL), project-based learning (POL), challenge-based learning, and active collaborative learning, among others. In all these methodologies the student interacts actively with his co-workers in small teams, and with his teacher, exchanging ideas and discussing progress in the solution or proposed solution to a specific scenario. The scenario or challenge should address a situation or problem as real as possible in order to better motivate students. In this sense, research-based learning can also be implemented seeking to have that active participation of students with their peers and teachers, discussing and analyzing scientific advances or proposing on their own contributions to the state of the art of specific disciplines. These

✉ Luis Neri
neri@tec.mx; lujanevi@gmail.com

Julieta Noguez
jnoguez@tec.mx

¹ Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey,
Calle del Puente 222, Col. Ejidos de Huipulco, Tlalpan,
14380 Mexico City, Mexico

activities can also and should be enriched with the interaction between teachers giving similar courses, thus promoting a fertile community of research-based learning, constituting an ideal community for the best learning of students and for their preparation as future researchers.

Therefore, an especially important issue is that teachers who intend to give courses in this modality (RBL) have to carry out a careful analysis while designing learning and teaching strategies to be sure that the research will actually be linked with the teaching. Even though they are professional researchers or may have experience with other inquiry-based-didactic techniques such as Problem-Based Learning (PBL) [2] or Project-Oriented Learning (POL) [3, 4], this care must be taken.

One important element to consider when designing RBL activities is the assessment of the initial student competencies at the start of the course. It is critical to know whether they have previously performed any kind of research, the research competencies they may already have developed and if they know how to access reliable information sources to perform their research.

For the first implementation phase of the RBL technique, the course on “Analysis and Science of Big Data Volumes” was selected, which is a topic taught in the 7th to 9th semesters in the Computational Systems Engineering career at Tecnológico de Monterrey, Mexico City Campus.

2 Theoretical framework

2.1 Research-based learning (RBL)

The growing importance of involving undergraduate students in professional research activities has been recognized in the university ecosystem since the end of the past century [5]. In their influential study, the Boyer Commission on Educating Undergraduates in the Research University has recommended that universities change a culture where students are passive receivers of information to one where they become active inquirers sharing a discovery adventure with staff, researchers and graduate students.

In fact, the Boyer report mentions that universities should develop research skills in their undergraduate students from the first year [5]. One of the main advantages would be to awaken students' interest in knowledge and the main problems that society faces in order that students may broaden their perspectives and focus their study areas. This report criticizes research universities where students are often disillusioned with the prestige of research leaders who are often not the ones who teach the classes, but, instead, their graduate assistants or other researchers of their work teams do it. It is emphasized that, nevertheless, there would be enormous value in involving students from the beginning of their

careers to acquire basic research skills, such as the search for information in authoritative and reliable sources, the critical analysis of knowledge and the development of oral and written communication skills to be able to present the results. All these competencies are very important for future professionals and can be taught to undergraduate students. Furthermore, students would feel better oriented in their chosen university careers, and, consequently, they could become more involved and motivated to carry out their studies [5].

From this perspective, RBL emerges as a learning methodology where research is regarded as an underpin to teaching at a wide range of levels. In addition to incorporating research outcomes, RBL develops students' awareness of processes and methods of inquiry and creates an inclusive culture of research involving staff and students [6]. In this enriching culture, undergraduate students can not only learn from the new discoveries and research methods but also engage themselves in the research processes themselves, thereby participating in the finding of new knowledge.

Due to its nature, RBL is related to other didactic techniques based on inquiry, such as Discovery-Based Learning [7], Inquiry-based Learning [8], Experiential Learning [9], Problem-Based Learning [2] and Project-Based Learning [3, 4]. The distinctive and common characteristic of these RBL techniques is that, in them, all activities are oriented to develop students' skills in research.

There are several approaches to implementing RBL in the classroom [10]. They include research-oriented (RO), research-based (RB), research-tutored (RT) and research-led (RL) (see Table 1). In RO, the curriculum emphasizes the understanding of processes by which knowledge is generated within the discipline. Students develop research and inquiry skills and techniques. In RB, the curriculum is designed with activities based on inquiry rather than on content acquisition of the subject; that is, students learn as researchers participating in research and inquiry activities. In RT, students learn through small group discussions with the teacher on research results. They participate in research discussions. Finally, in RL, the curriculum is structured around the subject content, which is determined by researchers and faculty. Teaching then occurs by transmission of information, and students learn through knowing about current research in the discipline. Some examples of activities for each of these approaches are given next [11].

Research-oriented (1) Invite students to a lab or a place to observe research in the real world; (2) make a presentation of research methods and approaches; (3) show experimental procedures and real exercises of computational issues in scientific disciplines; (4) ask students to read and perform searches from the bibliography of a research paper, analysing figures, diagrams, tables and simulators presented in the paper; (5) introduce students to a peer review of a research process (e.g., during the delivery of a scientific paper).

Table 1 RBL implementation approaches (taken from [11])

Student involvement	Emphasis on research content	Emphasis on research processes and problems
Students as Participants	Research-tutored: Course contents emphasizes students learning in small group discussions with a teacher about research findings	Research-based: Course content emphasizes students learning through inquiry-based, problem-based and project-based activities
Students as Audience	Research-led: Course content is based on disseminating factual and conceptual knowledge about the research interest(s) of the faculty member	Research-oriented: Course content emphasizes procedural knowledge about the research interest(s) of the faculty member or learning about the process by which knowledge is produced

Research-based (1) Introduce students to inquiry-based learning, where they are required to formulate and answer their own research questions; (2) ask students to make observations and formulate questions; then invite students to develop hypotheses individually or in teams and think of ways to test them; (3) ask students to perform research projects individually or in teams; (4) require students to publish a paper or produce a research result.

Research tutored (1) Assign graduate students as mentors to undergraduate students working on research projects; (2) assign undergraduate or graduate teaching assistants to support students in research skills for finding information in appropriate sources in libraries; (3) plan activities in which students interview faculty members about their research work.

Research-led (1) Explain the relevance of the current research in the course contents that are important to the teacher, faculty member or a given research group; (2) present research results in a given topic of the course; (3) invite a researcher to present relevant research for a course topic; (4) ask students to read specific research papers selected by the teacher.

2.2 Linking research to teaching in the undergraduate level

There are several different ways of linking research to teaching in the undergraduate level, the precise approach being tailored to the disciplinary context [10]. A framework for

Fig. 1 Framework for undergraduate research (taken from [1])

implementing research in undergraduate programs is proposed at the University of Alberta [1]. Their undergraduate research initiative considers a relationship between learning environment and learning outcomes (Fig. 1). In this scheme, student research practice can range from an instructor-centered environment to a learning-centered one. At the same time, learning outcomes can be developed in an increasing way, following Bloom's taxonomy scale.

Some successful examples of incorporating research in undergraduate programs in several institutions are briefly outlined next.

- *Warwick university* [12] In this University, research-based learning refers to all the dimensions of linking research and teaching, where research underpins teaching at a range of levels. It includes developing students' awareness of processes and methods of inquiry as well as incorporating outcomes of research into curricula, thereby creating an inclusive culture of research that involves staff and students.
- *Griffith* [13] Undergraduate students are better prepared for graduate level work through distinctive learning experiences in their programs, such as internationalization of the curriculum; research-based learning opportunities; and work placements that create learning situations which integrate academic theory with work practices.
- *University of California, Berkeley* [14] Berkeley is well recognized for its dynamic research environment. Undergraduate students have many opportunities to participate in this exciting research community. The interest of research pervades the classroom, where new knowledge and developments invigorate the learning process.
- *The University of Texas at Austin* [15] There is an annual celebration of undergraduate research and active creativity called, "Research Week," where undergraduates have great opportunities to explore and get involved in the vast and diverse community of research and creative work of this institution.

- *University of Leeds* [16] In this institution, the latest research contributes to the curricula of all programs, where students' independent research skills are actively developed. Students are provided with opportunities to put these skills into practice, so that, at the end of the program, they are able to undertake, with supervision, a final year project.
- *The Russell Group* [17] This group is comprised of 24 leading UK universities that are committed to maintaining excellence in research. In the Russell Group, outstanding teaching, learning experience and intensive research are integrated. Links with business and the public sector are established to have huge social, economic and cultural impacts across the UK and around the globe.

In this work, an approach was chosen that mainly focuses on the research-based model to emphasize research processes and problems.

2.3 Course-based undergraduate research experiences (CUREs)

In this section, a review of some relevant studies reported in the literature on Course-based Undergraduate Research Experiences (CUREs) is included. These works are mainly related to students enrolled in science, technology, engineering and mathematics (STEM) undergraduate programs.

Eagan et al. [18] suggest to federal and private institutions that they provide funding to STEM programs, in particular, to increase the numbers of underrepresented racial minority students in these programs. This recommendation is based on the fact that students in STEM undergraduate careers increasingly have shown intentions to enroll in graduate or professional school after having experiences in CUREs, and this would strengthen the graduate programs of universities. With a sample of more than 4000 aspiring student STEM majors, they use statistical techniques to examine how participation in undergraduate research affects STEM students' intentions to enroll in both STEM and non-STEM graduate and professional careers. Their study clearly indicates that participation in undergraduate research programs significantly impacted students' intentions and plans to enroll in STEM graduate programs. STEM aspirants who had undergraduate research experiences clearly showed a higher intention to pursue a graduate or professional degree in a STEM-related discipline compared to students who did not participate in the research programs. Important benefits and gains in terms of the production of domestic scientific expertise and the diversification of the scientific workforce are expected due to this.

Linn et al. [19] call attention to the fact that, although in most undergraduate research experiences students give high ratings to research experiences, in their opinion, the evidence for such claims is weak, because most studies rely

on student self-reporting surveys or interviews. They state that it is needed to undertake more rigorous research aimed to identify ways to design research experiences that promote integrated understanding. In this sense, these authors prompt the academic community to design powerful and generalizable assessments to document student progress and development of research competencies. That is, these authors suggest to carry out systematic, iterative studies to design research experiences that meet the needs of students, making effective use of resources and including multiple indicators of success. In addition to registering students' opinions about their undergraduate research experiences, Linn et al. [19] propose to analyse students' concrete research products (presentations and reports), to carry out direct measures of content learning gains as well as to study longitudinal evidence for persistence and to observe student activities.

Auchincloss et al. [20] provide a comparison of CUREs, on the one hand, and Research Internships (RIs), on the other hand, as two different models that involve undergraduate students in research activities. The main advantages of CUREs are: (1) by design, instructors in CUREs cover more students; that is, all students enrolled in the subject are covered; while, typically, in RIs, the instructor covers a small set of students; (2) in CUREs, one instructor can attend several students at the same time, while in the RI model, one instructor usually can attend only one student; and (3) CUREs are open to all students in a course, so that students invest time primarily in class (not outside), and they can also use a teaching lab instead of a faculty research lab, as in the case of RIs. On the other hand, CUREs also have several important disadvantages as compared to ISSs: (1) the amount of time spent on research activities in CUREs is typically smaller than for ISSs; (2) students in CUREs may be less interested in research as compared to students in ISSs, which can affect their motivation and that of their classmates; and (3) CURE students may not have access to sufficient mentorship or expertise to maximize scientific and learning outcomes as compared to students enrolled in ISSs.

From a review of relevant literature regarding CUREs, Corwin et al. [21] have determined some of the most relevant CUREs' outcomes. They have classified them in three categories: probable, possible and proposed:

- (i) Probable: Increased content knowledge, increased analytical skills, increased self-efficacy, external validation from a science community, persistence in science, increased technical skills, and career clarification.
- (ii) Possible: Increased project ownership, increased communication skills, increased motivation in science, increased collaboration skills, increased tolerance for obstacles, increased sense of belonging to a larger community, enhanced scientific identity and increased positive interactions with peers.

- (iii) Proposed: Increased access to faculty interactions, increased access to mentoring functions, enhanced understanding of the nature of science and development of self-authorship.

Thiry et al. [22] provide a comparison between novice and experienced undergraduate students' perceptions of their cognitive, personal and professional gains from engaging in scientific research in their education to become future scientists. They found that experienced students declare distinct personal, professional and cognitive outcomes when engaging in undergraduate research programs. Furthermore, these students also have a more sophisticated understanding of the methods of scientific research and higher intellectual skills, as well as they exhibit appropriate behaviors and temperament to become scientists. These results strengthen the assertion that undergraduate research experiences reinforce students' skills in research, and these improve with time.

Adedokun et al. [23] have studied the factors that determine the reported positive outcomes of undergraduate students enrolled in research experiences (undergraduate research experiences, or "UREs"). They are interested in investigating/exploring the logical and sequential relationships among three key outcomes of UREs; namely, research skills, research self-efficacy and aspiration for research careers. In their work, the logical relationship among outcomes, processes, contextual and participant factors that play in the undergraduate STEM research experiences is determined by student desire to persist in research careers. In particular, participant research self-efficacy (that is, a self-evaluation or belief about personal ability to take a course of action) and student research skills are the main factors that determine student aspirations to enroll in research careers.

In this context, participation in UREs enhance student career clarification, research, communications and critical thinking skills, the understanding of research processes and aspirations for graduate education and research careers in STEM disciplines. According to these authors, the most immediate and extensively studied outcome of UREs is, perhaps, their impact on student abilities and skills to perform research-related procedures, such as data entry, analysis and interpretation, as well as various laboratory skills and techniques.

Among the main positive impacts of URE experiences on student research, self-efficacy, according to Adedokun et al. [23], is the ability to think and work like scientists and to understand research processes; that is, the practical understanding of the nature of scientific knowledge and how science is conducted. Moreover, UREs promote student cognitive growth and professional socialization, and they improve their perceptions of self-efficacy and confidence to do research, so that they identify themselves as scientists or members of the scientific community. Finally, URE partici-

pants attend graduate school at much higher rates than their non-participant peers.

3 Goal

The main goal of this paper is to design and apply RBL activities in the modality of student participation in research tasks, so that students acquire research competencies in strategic exploration, such as inquiry and search and the elaboration of research papers.

4 Contributions

The most important elements of the contributions of this work are described below: (1) the methodology applied, taking into account the variability of the group's competencies and the heterogeneity of the students, because the course corresponds to a topic in which students from the seventh to the ninth semesters of a computer systems engineering career are included; (2) the instruments used to diagnose competencies; (3) the design and description of activities to promote the different competencies in research in phases; and (4) the design of the experiment based on the development of a research project in data science, in which students have to develop and apply data-analytics techniques and to obtain predictive models for patients with metabolic syndrome.

4.1 The methodology

A Research-based model was implemented to a group of undergraduate students of the Computational Systems Engineering career at Tecnológico de Monterrey, Mexico City Campus, enrolled in the "Analysis and science of big data volumes" course during the January-May 2018 semester. The semester includes 16 weeks. The objectives defined for the course were: (a) implement information-technology solutions within projects that involve the strategic analysis of data to generate value in the substantive tasks of a specific industry; and (b) redefine the business logic to guide it towards the analysis of data as a generator of differential value. To implement RBL successfully in this course, some research-based learning activities were designed to be carried out in parallel with the development of course contents, as shown in Fig. 2.

4.1.1 Diagnosis of RBL competencies

An important problem faced when designing the research activity was the heterogeneity of the course participants. Therefore, it was necessary to adapt a simplified instrument to determine the research competencies [24] that students

Fig. 2 Applied methodology

already have, such as information search strategies, tool management, databases and reliable information sources, research methodology, oral and written communication skills and abilities to do group work in research teams.

A question for each competence was designed in a 5-step Likert scale (from Total Agreement to Total Disagreement) to identify the student perception of their competencies. The results can be shown in Fig. 4.

4.1.2 Introductory workshop on research methodology

Based on these objectives and the fact that the group of students was very heterogeneous, an introductory workshop on Research Methodology was given that included the following topics: sources of information for research; aspects of evaluating Web sites as a research resource; use of the library, digital library and portals for research; tools for the prevention of plagiarism; and style formats in references and bibliography.

4.1.3 Monograph

To promote research-based learning, the first activity designed to develop basic research competencies was that students had to prepare a monograph on available Big Data used to develop Data Science and successful applications that have been reported in research works in the last 5 years. The activity was performed by teams of 5 students each. They considered the inquiry and documentary research of each platform, including where this had been applied. They had to make a comparative table and include conclusions.

Through this activity, it was expected that students would work on 3 essential issues described in the Framework for Information Literacy of the ACRL:

- (a) Information is built and contextual.
- (b) Research as Inquiry.
- (c) The search as strategic exploration.

It was explained to the students that a monograph is a document that deals with a particular topic, because it is dedicated to using diverse sources compiled and processed by one or several authors. It is a written, systematic and complete document with a specific or particular theme. The monograph includes detailed and exhaustive studies addressing various aspects and angles of the case, which must be extensively treated in depth. It follows a specific methodology and makes an important contribution, original and personal. The essential characteristic of a monograph is not its extension, but the character of work and quality; that is, the level of research.

Therefore, students were asked to prepare a monography on available Big Data platforms that support Data Science developments and on successful applications reported in research papers in the last 5 years. They were required to prepare a document with a description of each platform and containing the following elements:

- (a) Background of the platform.
- (b) Technical description.
- (c) Applications.
- (d) Trends.
- (e) Advantages and disadvantages.
- (f) Main conclusions.
- (g) References.

They also had to elaborate a comparative table of the surveyed platforms with a summary of their characteristics.

It was indicated that this activity would have a weight of 16% of the final grade of the course.

Students were also provided with a rubric that would allow them to cover the aspects to be evaluated during the preparation of the monograph. The corresponding performance criteria are shown in Table 2. The grading scale ranges from 1 to 100, where 70 is considered as the minimum acceptable grade.

The contents of the course were then delivered in a traditional model of teaching, and in week 12 of the semester, another research-based learning activity was assigned to prepare an article.

In parallel to the first research activities (Diagnosis of RBL competencies, Introduction to the research methodology, and Monography), the following subjects of the course were taught: Introduction to data science and the analysis of large volumes of data, Data analysis life cycle and Basic methods of data analysis using R. Then we started with the Research Project and with Advanced data analysis (theory, practice and technological tools), as shown in Fig. 3.

4.1.4 Approach to the challenge

Based on the diagnosis of the students' competencies, the introduction to the research workshop, the elaborated mono-

Table 2 Monograph evaluation rubric

Category	Very Good (91 to 100)	Good (81 to 90)	Fair (70 to 80)	Insufficient (less than 70)
Information search strategies (20%)	Adequate information search strategies were carried out in appropriate databases and reliable research sources	Appropriate information search strategies were carried out in some appropriate databases and reliable research sources	Information search strategies were carried out, but only a few were from appropriate databases and reliable research sources	Although information search strategies were carried out, they were not adequate and were based on inadequate databases and unreliable research sources
Information sources (15%)	All sources of information are recent, documented and in the requested format	All sources of information are recent and documented, but a few are not in the requested format	Most of the recent information sources are documented, but many are not in the requested format	Most sources of information are not recent and are not documented
Organization (15%)	The information is very well organized with well-written paragraphs and subtitles	The information is organized with well-written paragraphs	The information is organized, but the paragraphs are not professionally written	The information provided is not organized
Written communication (20%)	There are no errors in grammar, spelling or punctuation	There are no grammatical mistakes, but a few mistakes in spelling or punctuation	There are a few grammatical, spelling and punctuation mistakes	There are many errors in grammar, spelling or punctuation
Comparative Table (30%)	A good analysis and adequate comparison of the platforms was carried out	A good analysis was made, and most comparisons of the platforms were adequate	An analysis was made, but some of the comparisons of the platforms were not adequate	The analysis presented was not well done, and the comparisons of the platforms were not adequate

Fig. 3 Content of the course and the methodology applied

graph and the contents of the course, students were presented with a challenge to investigate trends and advanced techniques in Data Science, Big Data and Deep Learning in health; specifically, to obtain predictive models for metabolic syndrome and its surrogates in Mexico, based on open source databases.

4.1.5 Preparing a research project

To prepare students for the development of a research project, the following mini-workshops on research theory were interspersed in the theory of the course:

1. What is a research project?

It was explained to the students that the research considers the creative and systematic work to increase the knowledge of human beings, culture and society, as well as the use of this knowledge to design new applications.

It was clarified that a research project establishes or confirms facts, reaffirms the results of previous works, solves new or existing problems, supports theorems or develops new theories. It is thought that research projects can be used to develop greater knowledge about a topic. In this course, it is about promoting the research capacity of a student to prepare him or her for future jobs or reports.

Then, the characteristics of a research project were explained and the importance of having a written document that describes the activities to be carried out and the specifications and parameters of each of the activities.

2. Stages of a research process

In the logic of the research process, the students were explained the details of the following three basic stages:

- (a) *Stage of conception, planning and formulation of the scientific research project.* The student is invited to define what they want to investigate specifically in the topic of interest. It is suggested to ask in the form of a question in order to guide and specify what is going to be investigated and what is needed as an answer. The student is asked to express a hypothesis as a preliminary approximation that attempts to advance a theoretical explanation of the problem and thereby facilitate the solution. The student also explains the procedures, techniques, activities and other methodological strategies required for the investigation. As for students working in a team, planning is requested to define the process to follow in the collection of the information as well as the organization, systematization and analysis of the data.

- (b) *Stage of execution of the project or development of the investigation.* It must correspond to the type of study and its design based on the proposed objectives, the availability of resources, the type of study that will be carried out (exploratory, descriptive, experimental, observational, etc.), and it must offer details of its design (cohorts, cases and controls, clinical trials, etc.) and have a methodology to plan all the activities demanded by the project and to determine the human and financial resources required.
- (c) *Stage of preparation of the research report or communication of the results.* The results of the investigations are usually communicated publicly in different formats: (1) Reports of results; (2) Technical reports that contain preliminary working documents on the results of the research; (3) Communications and presentations at conferences to share with the academic community the results that can be submitted to criticism; (4) papers or scientific articles published in scientific journals that are reviewed and validated by external experts; (5) books and book chapters; and (6) other formats such as workshops and publication on the internet, etc.

3. Types of research

Every study must be understood as an exercise of measurement in each one of the sections of planning, execution and interpretation. It is therefore necessary to formulate objectives in a clear and quantitative way to make it very well known from the beginning what will be measured. If this first step is poor or unclear, the quality of the study will falter. The researcher must also have clarity about the level of scientific knowledge previously accumulated and developed by other researchers, as well as the unwritten information possessed by people who, by their stories, can help to gather and synthesize their experiences.

The following types of research and their characteristics were defined: Historical, Descriptive, Experimental, Quasi-experimental, Correlational and Ex post facto.

4. How to write a research article?

It was explained to the students that the writing process of a research article must be geared to publication in a research journal or participation in a contest. Therefore, the students need to consider:

- Identification of the readers: It is important to define what is the purpose of the document? Who will read it? How will the reader use it?
- Scope and size: The length, the level of detail and the style of the article must be decided.

- The concept: It was explained that in order to have good results, it is necessary to start with a plan and that the writers have different ways of developing plans.
- The outline: The sketch is the first draft. Place ideas on paper without worrying about style; make drafts of each section; develop the calculations; draw the figures; assemble references. All this gives freedom to shape a research article.
- Writing: The writing of each of the sections must be done with clarity, balance and legibility.
- The final product: In the final version of the article, care should be taken to have a good design, clear writing, appropriate headings and well-designed figures.

The importance of identifying a relevant journal for the publication of the project was explained. In addition, the main sections of a research article should include: (1) Introduction; (2) Theoretical framework; (3) Problem Statement; (4) Methodology; (5) Results and discussion; and (6) Conclusions.

4.1.6 Research project

The objective of the project was the design and evaluation of algorithms of Data Science, Big Data and Deep Learning to perform data analysis processing related to the metabolic syndrome and its surrogates in Mexico.

Metabolic syndrome (MS) is a cluster of conditions (increased blood pressure, high blood sugar, excess body fat around the waist and abnormal cholesterol or triglyceride levels) that, occurring together, increase the risk of heart disease, stroke and diabetes.

The prevalence of metabolic syndrome worldwide has increased dramatically in recent years, estimated at more than 25% of the adult population in most countries [25]. The prevalence of metabolic syndrome varies according to the classification used. The International Diabetes Federation is the one that classifies the largest number of individuals [26]. In Mexico, prevalence in the adult population has been reported in ranges from 13.6 to 59.9%, depending upon the diagnostic criteria used [27].

Health systems throughout the world are entering a new phase, where increasing amounts of massive, complex and multivariate data are obtained and stored on a daily basis throughout the life of a person. With robust, scalable and automated products, data science can significantly improve health outcomes through the use of specific probabilistic models for each patient. This is a field in which there is little existing research, and it promises to become a new industry supporting the next generation of health technology. The integration of data through spatial scales, from the molecular level to the population level, is one of the objectives of data science. A high degree of automation and, simultaneously,

a high degree of personalization are what is required for the informatics of future health and research.

Students were provided with an open source metabolic syndrome database [28], and each work team was asked to perform the following activities:

1. Perform the cycle of data analysis to obtain a clean data set; report the findings of the process (if there are missing data, if there are incomplete or out-of-range data); and propose a pre-processing to obtain predictive models.
2. It was emphasized to the students that the doctors wanted to obtain diagnostic and predictive models, and it was requested that the students make a first effort to build the initial data set.
3. The teams were asked to perform a search to obtain another set of metabolic syndrome data and use the best one supporting the decision.
4. They were asked to apply techniques to identify relevant variables and to reduce the dimensionality of the number of variables (for example, Principal Component Analysis).
5. The teams were asked to analyze the previous results and decide if it was necessary to apply MapReduce to extract the most relevant information from the data set.
6. The teams were asked to apply the different techniques seen in class to identify correlations, associations and predictive models on the chosen data set and to elaborate a proposal for a machine learning model.

With the previous activities, each student team was asked to:

- A. Prepare a research report.
- B. Prepare a research article and fill in the sections of a Congress template that was provided to them (minimum 6 pages, maximum 8 pages). The style details were provided in an additional document.
- C. Make a power point presentation to present the process, its results and conclusions.

During this period, the following topics of the course corresponding to Advanced Data Analysis (theory and practice with R libraries for advanced data analysis) were taught: (1) Clustering using the k-means method; (2) Rules of association; (3) Linear and logistic regression; (4) Simple Bayesian networks; (5) Decision trees; (6) Analysis of time series; (7) Text analysis; (8) Use of the MapReduce programming model for the analysis of semi-structured data; and (9) Data visualization techniques for the delivery of results.

Table 3 Questions of the questionnaire applied to identify the competencies of the students

Question number	Question
1	Through work I have undertaken in my previous courses, I have developed excellent information search strategies
2	I know and master a large amount of technological tools, databases and diverse sources to perform research tasks
3	I know, master and apply different research methodologies, according to the problem or type of problem to solve
4	I have excellent command of oral communication skills to report results
5	I have excellent command of written communication skills to report results
6	I have developed excellent skills to work in a research team to achieve successful results

5 Results

The results of the main activities of the research-based learning process are described next.

5.1 Diagnosis of RBL competencies

As mentioned in the Sect. 4.1.1 an important problem that was faced for the design of the research activity was the heterogeneity of the participants in the course, because the selected course is a topic for the career of Systems Engineers. It was necessary to adapt a simplified instrument to determine research competencies. Table 3 shows the questions of the questionnaire applied to identify the competencies of the students, according to their perceptions of themselves.

The answer to each question of competency was ranked by the student on a 5-step Likert scale (from Total Agreement to Total Disagreement) to identify their perceptions of their personal levels of acquired competencies.

The results obtained from the perception questionnaire of students on their previous research competencies are shown in Fig. 4.

Despite that more than 50% of the students responded that they have the basic research competencies (Total Agreement and Agreement), when given a review of basic concepts of the research process, it was identified that most students lacked those competencies.

5.2 Results of the monograph

The expected competencies from the monograph activity are that the student:

- Inquires about research sources and identifies important papers to elaborate a monography through an analysis of the papers; shows their contributions to the development of available state-of-the-art Data Science platforms.
- Performs critical analysis and elaborates a comparative table that includes the characteristics, advantages and disadvantages of Big Data.
- Discovers and gathers appropriate information about available Big Data platforms that would be useful in Data Science developments; proposes an adequate platform selection for a given posed problem.

According to the applied final assessment rubric shown in Table 2 regarding information search strategies, 100% of the teams carried out adequate strategies to search for information in databases and reliable sources of research. The sources of information cited by the students were current, documented and in the requested format. Regarding the organization of the written report, the information was very well organized with well-written paragraphs and subtitles. However, in the written communication, although most of the teams did not have grammar errors, some teams did have spelling or punctuation errors. Finally, most of the teams performed a good analysis, and most comparisons of the platforms were adequate.

Additionally, it was verified that no team had committed plagiarism, using the Safe Assign system of Blackboard. The range of matching found for the teams was between 3 and 14%, so it can be considered that the works were original.

5.3 Research project

The objective of the project was the design and evaluation of algorithms of Data Science, Big Data and Deep Learning to perform data analysis processing related to metabolic syndrome and its surrogates in Mexico, as mentioned previously.

Once the project objectives were established, some students stated that “from the beginning of the project, the fact that we were dealing with a topic that is still controversial emphasizes the importance of creating a model that allows an accurate prediction with minimal error, which could not happen otherwise”.

All the teams applied the cycle of data analysis, performing pre-processing of the data to obtain a clean data set. Because a very large database with many variables was provided, students were asked also to apply techniques, such as principal component analysis (PCA), to identify relevant variables and reduce the dimensionality of the number of variables.

Next, the teams were asked to analyze the previous results and decide if it was necessary to apply MapReduce to extract the most relevant information from the data set. All the teams chose to apply this technique and carried it out successfully.

Fig. 4 Self-diagnosis of initial research competencies

In accordance with the requirements of the doctors who requested the project, the teams were asked to apply the different machine learning techniques taught in class to identify correlations, associations and predictive models in the clean data set and to prepare a proposal for an automatic learning model.

All the teams of students reached the goal of the project, because they managed to obtain a coherent and accurate predictive model with respect to the problem that had been raised regarding Metabolic Syndrome.

5.4 Technical project report

Most of the teams delivered their well-organized technical project report with the proper description of the techniques used and the correct code, and they added conclusions and future work. The technical report of the project, requested from each team of students, contained the detail of the following sections:

5.4.1 Cover

The cover of the technical project report must contain the name of the institution, the course name, the project title, the names of the students of the team, and the delivery date.

5.4.2 Database

Description of the Open Source Database, obtained from Fernando García and Ricardo Salcedo <https://data.world/yehter/metabolic-syndrome-parameters-nhanes-2011-2012>

5.4.3 Life cycle of data analysis

Description of the data analysis cycle application:

- Description of the establishment of the project objective.
- Data cleaning: The students analyzed the columns of the data set to perform the process of cleaning the data.
- Data collection and management: The students searched databases with information relevant to the metabolic syndrome in Mexico and found some alternative databases that consider different relevant factors related to the metabolic syndrome and the fact of suffering or not the metabolic syndrome.

5.4.4 Application of machine learning techniques

- Principal component analysis: The students applied techniques to determine the most relevant variables in the diagnosis of metabolic syndrome; applied a dispersion diagram of the first variables; determined the levels of correlation among the different columns and established criteria to reduce the data set and, thus, were able to perform the operations to continue working with the data set.
- MapReduce: Because the information was in a table format (data-set), the students had to make a transformation to apply the Map and Reduce techniques, so that everything was first transformed into a plain text and then into a character format to later pass it to a list and, finally, to a Dlist. The students investigated to find a function that obtained the words from the Dlist, and they created the function, “splitfunc.” In this process, they determined that, because too much information was lost, they returned to the process data set of a).
- Logistic regression model: It was decided to use a logistic regression model, because the dependent variable was binary or dichotomous and was adequate for the problem that was being addressed. Subsequently, the data was divided into 2 groups; namely, a training set to

obtain the predictive model and another set to test the model.

5.4.5 Conclusions

All the teams of students considered that the objective of the project was fulfilled, because they managed to obtain a coherent and precise predictive model with respect to the problem that had been posed to them regarding Metabolic Syndrome.

5.5 Research paper

To prepare a research article, students were provided with a template for the Congress with sections: (1) Introduction; (2) Theoretical Framework; (3) Problem Statement; (4) Methodology; (5) Results and Discussion; and (6) Conclusions. The range of the requested article was a minimum of 6 pages and maximum of 8 pages. The details of the style were provided in a complementary document. In general, the results of the teams were satisfactory.

The six teams of students each managed to write a quality article that could be sent to a research conference in the area. In spite of the little time that was had for the elaboration of the articles of investigation, all the teams fulfilled the requested sections. In particular, they adequately presented the objectives, methodology, results, discussion and the conclusions.

Some teams emphasized the usefulness of the linear regression model, which gave them information about the data that was used for the prediction of objective variables, and the coefficients showed that the model was relatively accurate. They also stated that Data Science was a great tool to use to perform the analysis of the data and then create the models.

5.6 Final presentation to present the process, the results and the main conclusions

For the final presentation of the research projects, a group of professors and doctors was invited as synods. The jury was composed of four synods. The evaluation of the teams was carried out on a 5-level Likert scale (1 = Very good, 2 = Good, 3 = Fair, 4 = Bad and 5 = Very bad). In general, the teams were well evaluated. Of the 6 teams (made up of 5 students each), 2 teams obtained qualification 1 (Very good); 2 teams obtained the qualification of 2 (Good); and 2 teams obtained the qualification of 3 (Fair). In general, the synods' observations were directed towards pointing out opportunities for improvement in the presentation of the results and the conclusions of each team. Additionally, it was suggested to the teams to explore an additional model or to add more variables to improve the quality of their model.

6 Conclusions

New engineers must be able to propose and implement information technology solutions to projects involving strategic data analyses in order to generate value to essential tasks of a specific industry. As well, they must be able to redefine the business logic to be oriented to data analysis as a differential value generator.

In this proposal, the student formation is enriched by acquiring research competencies to apply successful strategies of inquiry, investigation and analysis of information, taking advantage of resources of data bases and reliable information sources.

This study shows the importance of performing an initial diagnosis to determine the students' research skills. Moreover, it is necessary to train them during the semester to acquire and develop the research skills. Students were able to develop these skills through a research project that involved the design and evaluation of algorithms of Data Science, Big Data and Deep Learning to perform data-analysis-processing related to metabolic syndrome and its surrogates in Mexico. They wrote a monograph where they developed skills to search information in appropriate, validated sources. They also developed skills to analyze data using statistical methods. Finally, they were able to design the sketch of a research paper and present the results to a jury.

The data analysis proposed in the selected course is not about numbers, but about asking questions, developing explanations and testing hypotheses. The foregoing provides a large area of opportunity for a student in Engineers in Computational Technologies to perform data science, and, for this, the student must have, among others, the following research competencies:

- Collects, processes and extracts value from the diverse and extensive databases.
- Applies imagination to understand, visualize and communicate findings from non-scientific data.
- Develops the ability to create data-driven solutions that increase benefits and reduce costs.

Data scientists can work in all industries and deal with large data projects at all levels. The proposed activity will allow the acquisition of the aforementioned competencies.

This study supports the assertion that it is very important to start developing RBL competencies in undergraduate engineering students early in order to prepare them for graduate studies or their professional lives.

7 Release and future testing

The authors are exploring the use of the RBL methodology in the first third of engineering careers in basic courses, such as Physics or Mathematics, so that students would reach their final semesters with some basic training in research competencies, which would then be reinforced and widened at in the final semesters of their careers (curricula).

Acknowledgements The authors would like to acknowledge the financial and the technical support of Writing Lab, TecLabs, Tecnológico de Monterrey, Mexico, in the production of this work.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- University of Alberta (adapted from the Committee on the Learning Environment (CLE) sub-committee report: Teaching, Research and Discovery Learning: Recommendations for a great University., January 2010). <http://www.uri.ualberta.ca/DefiningUndergraduateResearch.aspx>. Accessed 8 Apr 2018
- Hung, W., Jonassen, D.H., Liu, R.: Problem-based learning. *Handb. Res. Educ. Commun. Technol.* **3**, 485–506 (2008)
- Dutson, A.J., Todd, R.H., Magleby, S.P., Sorensen, C.D.: A review of literature on teaching engineering design through project-oriented capstone courses. *J. Eng. Educ.* **86**(1), 17–28 (1997)
- Mills, J.E., Treagust, D.F.: Engineering education—is problem-based or project-based learning the answer? *Aust. J. Eng. Educ.* **3**(2), 2–16 (2003)
- Boyer, E.L.: The Boyer Commission on Educating Undergraduates in the Research University, Reinventing Undergraduate Education: A Blueprint for America's Research Universities, p. 46. Stony Brook, NY (1998)
- Warwick Learning and Development Centre. England, UK. <https://warwick.ac.uk/services/ldc/resource/rbl/whatis>. Accessed 12 April 2018 (2018)
- Bruner, J.S.: Toward a Theory of Instruction, vol. 59. Harvard University Press, Cambridge (1966)
- Garrison, D.R., Arbaugh, J.B.: Researching the Community of Inquiry Framework: Review, Issues, and Future Directions. *Internet Higher Educ.* **10**(3), 157–172 (2007)
- Beard, C.M., Wilson, J.P.: *Experiential Learning: A Best Practice Handbook for Educators and Trainers*. Kogan Page Publishers, London (2006)
- Griffiths, R.: Knowledge production and the research-teaching nexus: the case of the built environment disciplines. *Stud. High. Educ.* **29**(6), 709–726 (2004)
- Center for Teaching Excellence. University of South Carolina. Retrieved August 13th, 2018, from https://sc.edu/about/offices_and_divisions/cte/teaching_resources/maintainingbalance/link_teaching_research/
- The University of Warwick, UK. <https://warwick.ac.uk>. Accessed 12 May 2018 (2018)
- Griffith University, Australia. <https://www.griffith.edu.au/research>. Accessed 12 May 2018 (2018)
- University of California, Berkeley, USA. <http://research.berkeley.edu/>. Accessed 12 May 2018
- The University of Texas at Austin, USA. <https://ugs.utexas.edu/researchweek>. Accessed 12 May 2018
- University of Leeds, UK. http://ses.leeds.ac.uk/info/22222/leeds_curriculum/1125/enhancing_teaching_and_research-based_learning. Accessed 12 June 2018
- The Russell Group, London, UK. <http://www.russellgroup.ac.uk/about/>. Accessed 12 June 2018
- Eagan Jr., M.K., Hurtado, S., Chang, M.J., Garcia, G.A., Herrera, F.A., Garibay, J.C.: Making a difference in science education: the impact of undergraduate research programs. *Am. Educ. Res. J.* **50**(4), 683–713 (2013)
- Linn, M.C., Palmer, E., Baranger, A., Gerard, E., Stone, E.: Undergraduate research experiences: impacts and opportunities. *Science* **347**(6222), 1261757 (2015)
- Auchincloss, L.C., Laursen, S.L., Branchaw, J.L., Eagan, K., Graham, M., Hanauer, D.I., Towns, M.: Assessment of course-based undergraduate research experiences: a meeting report. *CBE Life Sci. Educ.* **13**(1), 29–40 (2014)
- Corwin, L.A., Graham, M.J., Dolan, E.L.: Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. *CBE Life Sci. Educ.* **14**(1), 1 (2015)
- Thiry, H., Weston, T.J., Laursen, S.L., Hunter, A.B.: The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. *CBE Life Sci. Educ.* **11**(3), 260–272 (2012)
- Adedokun, O.A., Bessenbacher, A.B., Parker, L.C., Kirkham, L.L., Burgess, W.D.: Research skills and STEM undergraduate research students' aspirations for research careers: mediating effects of research self-efficacy. *J. Res. Sci. Teach.* **50**(8), 940–951 (2013)
- Cañal, P.: ¿Cómo evaluar la competencia científica? Capítulo 11 del libro de Pedrinaci, Caamaño, Cañal y de Pro “El desarrollo de la competencia científica: La evaluación de la competencia científica requiere nuevas formas de evaluar los aprendizajes”. Graó, Barcelona (2012)
- Khan, R.J., Gebreab, S.Y., Sims, M., Riestra, P., Xu, R., Davis, S.K.: Prevalence, associated factors and heritability of metabolic syndrome and its individual components in *African Americans: the Jackson Heart Study*. *BMJ Open* **5**(10), e008675 (2015)
- Guo, H., Liu, J., Zhang, J., Ma, R., Ding, Y., Zhang, M., Mu, L.: The prevalence of metabolic syndrome using three different diagnostic criteria among low earning Nomadic Kazakhs in the Far Northwest of China: new cut-off points of waist circumference to diagnose MetS and its implications. *PLoS ONE* **11**(2), e0148976 (2016)
- Aguilar-Salinas, C.A., Rojas, R., Gómez-Pérez, F.J., Valles, V., Ríos-Torres, J.M., Franco, A., Sepúlveda, J.: High prevalence of metabolic syndrome in Mexico. *Arch. Med. Res.* **35**(1), 76–81 (2004)
- Dataworld. <https://data.world/yehter/metabolic-syndrome-parameters-nhanes-2011-2012>. Accessed 12 Jan 2018 (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.