

From Methodological Review to Research Design: Decisions Adopted in an Educational Robotics Study for Physics Teaching in a Rural Context

Luis Augusto Madera Angulo

Politécnico Colombiano Jaime Isaza Cadavid, Medellín, Antioquia, Colombia.

luismadera@elpoli.edu.co

<https://orcid.org/586-603-586-603>

HOW TO CITE:

Luis Augusto Madera Angulo
(2026). From Methodological
Review to Research Design:
Decisions Adopted in an
Educational Robotics Study for
Physics Teaching in a Rural
Context.
International Journal of Special
Education, 41(16s), 586-603

ABSTRACT:

Methodological decisions in educational research should not be presented as isolated elements or as a list taken from research-methods textbooks. This article reconstructs the process through which a methodological review of educational robotics studies informed the methodology of a doctoral investigation conducted with upper-secondary students at a rural school in Apartadó, Colombia. The final version of the dissertation was treated as the main source, and the documentary corpus used during its development was examined. The original Scopus search retrieved 184 records published between 2018 and 2023, from which 50 documents were selected according to relevance and citation count and then the search is broadened, selecting 21 articles published between 2024 and 2026. A broader group of later studies and specific literature on the CLASS and PMQ/SMQ-II instruments were also available for comparison. The analysis showed a wide variety of approaches in educational robotics, limited use of pretest-posttest studies with a control group, frequent reliance on questionnaires, interviews, and observations, and the need to distinguish academic performance from attitudes and motivation. These findings were considered together with the research question, the characteristics of intact school groups, the institutional conditions, and the availability of measurement instruments. As a result, the study adopted a positivist paradigm, a quantitative method, a quasi-experimental design with experimental and control groups, a correlational component, and an analysis plan suited to the nature of the data. The article argues that a methodological review can be used to justify the connected decisions that make up a study rather than merely to summarize previous research.

Keywords: methodological review; methodological decisions; educational robotics; physics teaching; rural education; validity and reliability; quasi-experimental design.

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.
Open access publication under the
terms and conditions
of the Creative Commons
Attribution (CC BY)
license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Methodology is a fundamental part of any research, because it defines the way in which the necessary information will be obtained to answer

the question posed. Despite this, in some works it is presented as a section that is constructed after having formulated the problem and objectives, almost as if the researcher only had to

select an approach, a design and several instruments from a previously established list. In reality, these decisions are related to each other. The choice of a variable conditions the instrument; the type of instrument limits or expands the possible analyses; the design determines the level of comparison that can be made; and the conditions of the context may make a methodology that, in theoretical terms, seems adequate, unviable.

Literature review is usually used to recognize the progress of a topic, describe results and build the theoretical framework. In the process of preparing the thesis analyzed in this article, it was also used for another purpose. The documents were reviewed to learn how other researchers had conducted their studies: what approaches they used, which groups participated, how they collected the information, what instruments they selected, what analysis they applied, and what difficulties they encountered during experiences with educational robotics. In this way, each article provided not only conclusions on robotics, but also concrete examples of methodological decisions.

The doctoral research was carried out in a rural educational institution in the municipality of Apartadó, in the department of Antioquia. The problem was related to the low academic performance in physics and the presence of difficulties in the attitude and motivation of high school students. The interest consisted in evaluating a didactic strategy supported by educational robotics and not simply in describing the acceptance of a technology. This involved measuring the initial state of the students, implementing a proposal, making a new measurement and comparing the results between two groups. It was also necessary to determine whether attitude and motivation were related to academic performance.

When the methodological review began, the literature on educational robotics showed very different experiences. Some research focused on computational thinking, others on programming, problem solving, collaboration, identity, interest, motivation, or performance. We also found

reviews, case studies, qualitative papers, mixed research and quantitative studies. This variety was useful because it allowed us to recognize numerous possibilities, but at the same time it made it difficult to translate a complete methodological model from a single article. The populations, the duration of the interventions, the school content and the technological conditions were not the same.

Reviews by Anwar et al. (2019), Ioannou and Makridou (2018), and Pedersen et al. (2020) showed the breadth of the field and the diversity of outcomes assessed. In the empirical studies, experiences were found in basic, secondary and university education, as well as curricular and extracurricular activities. Studies such as those by Kert et al. (2020) and Lopez-Caudana et al. (2020) made it possible to observe quantitative and comparative procedures, while other studies delved into the experience of participants through interviews, observations, or interaction analysis. Therefore, the methodological decision could not consist of choosing the procedure that appeared most often. It was necessary to analyze which of these possibilities best fit the question and the context of the thesis.

The selection of instruments also required a review of its own. To measure attitude and beliefs towards physics, the Colorado Learning Attitudes about Science Survey (CLASS) was considered. This questionnaire had extensive initial validation, but subsequent research indicated that its structure could behave differently in different populations (Adams et al., 2006; Badeo & Duque, 2023; Kaltakci-Gurel, 2021; Kontro & Buschhüter, 2020). For motivation, the evolution of the Science Motivation Questionnaire was reviewed, from its initial version to the SMQ-II and the specific adaptations to physics (Glynn et al., 2009, 2011). An important insight emerged from this review: it was not enough to say that an instrument had been validated. It was necessary to know what type of evidence had been obtained, with what population and what the behavior of the instrument was in the own sample.

This article answers the following question: how did the methodological review of research on educational robotics and psychometric studies support the decisions that shaped the methodology of a doctoral research on the teaching of physics in a rural context? The objective is to reconstruct the process followed to move from the information found in the literature to the definition of the paradigm, the approach, the design, the sample, the instruments, the procedure and the statistical analyses used.

The article does not seek to establish a single valid route for all educational robotics research. Its purpose is to show, from a specific case, how the relationship between the reviewed evidence and the final decisions can be made visible. This reconstruction allows us to understand that the methodology was not formed through independent elections. It developed gradually, by comparing what other researchers had done with what the research problem demanded and with what was possible to do in the educational institution.

2. Methodology of the methodological review

2.1. Type of work and sources used

This article corresponds to a documentary reconstruction of the methodological process followed during the thesis. The expression methodological review is used because the main interest was in examining the procedures used by other research and the way in which these procedures served to define the study itself. It is not called a systematic review, because the original process did not include a registered protocol, a formal risk of bias assessment, or all the components that are currently required to use that designation.

The final version of the doctoral thesis, presented in April 2026, was taken as the main document. The research question, objectives, paradigm, method, design, population, sample, instruments, application procedure, statistical analyses, and ethical considerations were extracted from it. This rule avoided mixing preliminary decisions with those that were finally applied. During the development of a thesis, the variables, the

population or even the form of intervention may change; For this reason, previous versions were not used to correct or replace the final methodology.

The second group of sources was made up of articles on educational robotics. For the historical review that accompanied the thesis, 50 documents were used. In the preparation of this article, 71 documents on the subject were available, but the additional 21 were used only as support to contrast trends and not to affirm that they influenced decisions made before their publication or incorporation into the corpus.

The third group corresponded to studies on attitude and motivation instruments. Documents related to CLASS, its adaptations and psychometric analyses were reviewed, as well as research on the Science Motivation Questionnaire, the SMQ-II and its versions applied to physics. These studies were not treated as general background to robotics, but as specific evidence for the selection and use of questionnaires.

2.2. Search and selection of educational robotics articles

The original search was performed in Scopus with the equation:

("educational robotics" OR "school robotics") AND ("secondary education" OR "middle school" OR "High school" OR "K-12").

Documents published between 2018 and 2023 were considered. The consultation retrieved 184 records. From this set, 50 documents were selected, taking into account the number of citations and the direct relationship with the problem of the thesis. The number of citations was used as a practical criterion to reduce the volume of documents, but not as an automatic proof of quality. It was also reviewed that the studies provided useful information on educational robotics, school population, research methodologies, instruments, results or implementation conditions.

Works that mentioned robotics in a secondary way, research belonging to fields other than the object of study and documents in which the

robot functioned mainly as a tutor or social agent were excluded, because the thesis was oriented to robotics as a tool for construction, programming and learning. Lower priority was also given to research carried out exclusively in primary education when its procedures were not comparable with secondary education.

The selection made it possible to handle a corpus large enough to recognize regularities. However, it also generated limitations. More recent articles had less time to accumulate citations, and searching a single database could have left out regional publications. These conditions were preserved in the reconstruction, because they are part of the procedure actually performed.

2.3. Organisation of information

The information from the documents was taken to a matrix. For each study, the purpose, population, educational level, focus or type of research, collection techniques, instruments, procedure, resources used, pedagogical theories, teaching methods, limitations, and recommendations were recorded, where available. In the review phase, text mining and artificial intelligence tools were used to support the classification and synthesis of an extensive corpus. However, the information used to write the thesis and this article was contrasted with the original documents.

The matrix made it possible to separate aspects that could be confused in a general reading. For example, a survey can be a collection technique and the questionnaire its instrument; a study may be declared mixed and still give greater weight to qualitative data; One intervention may use the same technological device as another, but differ completely in duration, objectives and way of working. The organization by fields made it easier to compare these elements without depending on the order in which the documents were read.

For CLASS and PMQ/SMQ-II a different reading was created. The number of items, the dimensions, the population in which they were applied, the translation or adaptation procedures, the factor analyses, the internal consistency coefficients, the test-retest stability and other available evidence were recorded. The difference

between validity and reliability was maintained. A Cronbach's alpha was interpreted as evidence of internal consistency and not as a complete demonstration of validity.

2.4. Form of analysis

The analysis was carried out in two moments. First, trends in the corpus of educational robotics were identified. Frequencies were calculated on types of research, techniques, instruments, pedagogical foundations, teaching methods and technologies, according to the classification stated in the thesis. Second, these trends were compared with the needs of doctoral research.

The comparison was made using a simple sequence. The alternative found in the literature was identified, what information it could provide was examined, its limitations were reviewed, and it was contrasted with the question, the variables, the population, and the institutional conditions. Based on this analysis, the decision adopted was justified. Therefore, the frequency of a method was a reference point, but not the sole criterion for selecting it.

3. Results of the methodological review

3.1. Approaches and designs found

The classification of the 50 documents showed that educational robotics has been studied from very diverse perspectives. 16% corresponded to systematic reviews and 10% to descriptive works without direct participation of students or teachers. Among the studies carried out with participants, 38% were classified as mixed, 20% as qualitative and 16% as quantitative.

The presence of mixed studies is related to the number of aspects that are usually observed in a robotics experience. Researchers are not limited to academic performance. They also study collaboration, interest, attitudes, identity, programming skills, problem solving, and perception of teachers and students. To cover several of these aspects, it is common to combine questionnaires with interviews, observations, products prepared by the participants or class records.

Qualitative studies made it possible to understand processes that a test does not easily show. In robotics teams, for example, leadership positions, women's participation, power relations, and identity have been analyzed (Hennessy Elliott, 2020; Sullivan & Bers, 2019). The mixed studies combined performance measures with information from participants' experience. This variety was valuable to get to know the field, but not all the options answered the thesis question in the same way.

In the research techniques, the interview appeared in 28% of the documents and observation in 22%. The survey accounted for 14%, content analysis 12% and the case study 6%. Among the instruments, the questionnaire was the most frequent, with 20%. Pretest-posttest tests with a control group were found in 8% of the documents. Studies with post-test without a control group, diaries, academic records and recordings had lower percentages.

This result had a direct impact. The thesis sought to evaluate changes and compare a group that would receive the strategy with another that would continue with the usual process. Interviews and observations could provide important information, but they were not sufficient to answer a question related to effect, differences and associations between quantified variables on their own. The low presence of controlled designs did not lead to discarding them; on the contrary, it allowed us to recognize that they could provide a type of evidence less frequent in the corpus.

3.2. Pedagogical trends and definition of the intervention

In the review, a wide presence of constructivism and constructionism was observed. According to the codification carried out, constructivism was associated with 88% of the documents and constructionism with 50%. These figures do not mean that all the articles extensively developed a pedagogical theory, but they do show that robotics was understood, in many cases, as an activity where the student learns by building, programming, testing and correcting.

Project-based learning appeared in 90% of the documents. Collaborative learning was present in 54% and practical learning in 46%. Gamification, inquiry, flipped classroom, active learning and other proposals were also found, although less frequently. Research such as that by Kucuk and Sisman (2018), Sisman and Kucuk (2019), and Ioannou et al. (2018) shows that robotic activity is often organized around challenges, projects, and products that must work.

This trend helped to specify the independent variable. It was not methodologically correct to define it only as the use of educational robotics. A board, a robot or a program is not in itself a didactic strategy. It was necessary to establish how the sessions would be organized, what activities the students would carry out, what physics content would be worked on, and what role the teacher would have. Consequently, the intervention was structured using the ASSURE model and constructionism, project-based learning, collaborative work and practical activities were combined.

The pedagogical review did not automatically determine the design of the strategy. Pre-test results and knowledge of the context were also considered. The students had little experience in programming and electronics, so a progressive sequence was designed. First circuits were addressed, then programming with Micro:bit and finally construction and control of robots. This sequence responded both to the literature and to the characteristics observed in the group.

3.3. Technological resources and context conditions

The documents reviewed used a wide range of resources. Commercial platforms, LEGO, Arduino, Micro:bit, Raspberry Pi, simulators, block languages, drones, computer vision and other solutions were found. In the corpus classification, open or low-cost options appeared in 24% of the documents; platforms and kits of different brands by 22%; LEGO and block-based programs by 20%; simulators by 14%; Arduino at 8% and ROS at 4%.

The variety of tools indicated that there was no single platform that could be considered the best

for all cases. Costelha and Neves (2018) point out precisely the difficulty that teachers face in choosing among many available resources. Other studies show virtual or simulated alternatives to reduce costs and facilitate access (Camargo et al., 2021; Dietz et al., 2022). In rural contexts, Manikutty et al. (2019) show the importance of considering infrastructure, teacher training, and availability of materials.

The institution where the thesis was carried out had Micro:bit cards and laptops, although internet connection was limited. For this reason, resources were chosen that could be used without relying on permanent connectivity. The selection of Micro:bit and MakeCode was not made because these resources appeared in the largest number of articles, but because they were available, compatible with the level of the students and could be integrated with low-cost components.

This decision shows the relationship between review and context. The literature allowed us to know common possibilities and problems; The institution established the real limits. It would have been of little use to select a system with better technical characteristics if it could not be sustained during the intervention or if it required a non-existent infrastructure.

3.4. Aspects insufficiently addressed in the literature

The review grouped 29 recommendations or outstanding issues into 12 themes. Among the most frequent were the effective integration of robotics in the classroom, longer studies and

evaluation in different contexts. Access to low-cost platforms, cultural and geographical factors, teacher training, comparison of methods and research in rural areas also appeared.

These aspects were not taken as a list of topics that should all be included in the thesis. They were used to assess the relevance of some decisions. The research would be carried out in a rural context with connectivity limitations and with school groups already formed. In addition, it intended to make comparisons before and after the intervention. Therefore, several issues reported in the literature coincided with actual conditions of the study.

Rurality was not treated only as a geographical location. It influenced the availability of resources, the way work was organized, the possibility of maintaining the intervention and the need to use accessible tools. Similarly, the scarce presence of comparative designs in the corpus helped to justify the use of a control group, although the school organization did not allow students to be assigned individually at random.

The literature after the original period has continued to examine these questions. Works have been published on rural teacher leadership, accessible platforms and measurement of participation in robotics. These studies were not used to claim that they had determined the methodology of the thesis, but they do allow us to verify that the decisions made are still relevant to the field.

Table 1. Methodological trends and relationship with the design of the thesis

Dimension	Finding of the review	Observed Options	Use in doctoral research
Type of research	Mixed studies were the most frequent; there were also qualitative, quantitative, reviews and descriptive	Case studies, surveys, mixed designs, comparisons and reviews	A quantitative method was selected because the question required measuring changes and associations
Comparison of results	Pretest-posttest designs with a control group had low presence	Single Post-Test, Observation, Interviews, Records and Comparison Groups	A quasi-experimental design was adopted with initial and final measurement

Dimension	Finding of the review	Observed Options	Use in doctoral research
Techniques and instruments	Interviews and observations were frequent; The questionnaire was the most used instrument	Quizzes, interviews, observation, videos and products	Standardized instruments were chosen for the defined variables
Intervention	Predominance of project construction, collaboration and practice	Projects, gamification, inquiry, flipped classroom and extracurricular activities	The strategy was organized with ASSURE, constructionism, projects, collaborative work and practice
Resources	A wide variety of devices and programs existed	LEGO, Arduino, Micro:bit, simulators, drones and block languages	Micro:bit and MakeCode were chosen for availability and possibility of offline use
Background	It was recommended to expand the research to diverse realities and with less access	Urban, rural, university, school and extracurricular contexts	The study was carried out in a rural institution and its conditions were described

4. From methodological review to research decisions

4.1. Paradigm and method

The thesis was located in the positivist paradigm and used the quantitative method. This decision was not made because quantitative studies were the most numerous. In the reviewed corpus, they were less present than mixed and qualitative ones. The choice was related to the type of response sought.

The main variables could be expressed through scores. Academic performance was measured based on correct answers in a physics test; attitude and motivation were obtained using questionnaires with a Likert scale. The question required comparing groups, analyzing changes and studying relationships between variables. These operations required measurement, data organization, and statistical contrast.

The quantitative method did not imply ignoring the complexity of attitude or motivation. It was recognized that these constructs are not exhausted in a score. However, for the objectives set, it was necessary to define indicators and use instruments that would allow results to be compared. Qualitative research would have been appropriate to understand particular experiences, relationships within teams, or meanings attributed to robotics, but it would have answered different questions.

The paradigm and method were then defined based on three elements: the form of the question, the operational nature of the variables, and the type of analysis required. The review helped to learn about alternatives, but the final decision was made by checking their correspondence with the study.

4.2. Quasi-experimental design

The research used a quasi-experimental design with pre-test and post-test, experimental group and control group. This decision was based on the need to have an initial and a final measurement. Without the pre-test, it would have been difficult to establish whether the groups presented comparable conditions before the intervention. Without a control group, any changes could be confused with the passage of time, usual school activities, or other factors.

The institution worked with courses already formed. It was not feasible to reorganize students individually to create new groups. For this reason, an experiment with individual randomization was not proposed. Intact groups were selected, a condition typical of a quasi-experimental design.

The population was made up of 206 tenth and eleventh grade students: 120 tenth graders and 86 eleventh grade students, according to enrollment information. For the intervention, the eleventh groups were chosen by non-probabilistic convenience sampling. This choice was related to the students' trajectory in physics and the feasibility of implementing the proposal in the time available.

A draw was held between the three groups of eleventh to define their function. Group 11A was assigned as experimental, 11B as a control, and 11C did not participate. The final sample was made up of 56 students, 27 in the experimental group and 29 in the control group. The lottery narrowed the investigator's direct decision about which course would receive the intervention, but it does not turn the study into an individual randomized trial.

Students who agreed to participate and attended at least 80% of the sessions were included. Those who did not meet the minimum attendance, withdrew or decided not to continue were excluded. These conditions were necessary to ensure sufficient exposure to the strategy, without modifying the normal rules of school organization.

4.3. Correlational Scope and Group Comparison

In addition to comparing the results of the intervention, the research sought to examine the relationship between attitude, motivation, and academic performance. The review showed that socio-affective variables occupy an important place in science and robotics teaching studies. However, motivation, interest, attitude and self-efficacy are not always clearly differentiated.

In the thesis it was decided to measure attitude and motivation using different instruments. The attitude was understood from the beliefs about physics and its learning; Motivation was organized into dimensions such as intrinsic motivation, career, self-determination, self-efficacy, and grades. Academic performance was treated as another variable, measured by a physics test.

The correlational component allowed us to analyze whether students with better scores in attitude or motivation also presented higher performance. This analysis was not used to assert causation. A correlation indicates association, but does not demonstrate that one variable directly produces changes in another. The comparison between groups and the correlational analysis were, therefore, two different components of the study.

4.4. The rural context in methodological decisions

The educational institution is located in a rural environment of Apartadó. This condition influenced everything from the formulation of the problem to the selection of the technology and the application procedure. The literature indicated that educational robotics experiences cannot be transferred without adjustments between contexts with different resources, connectivity and teacher training.

In the institution there were 75 Micro:bit cards and approximately 225 laptops. At the same time, the internet connection was not stable. This situation showed that the problem was not simply the absence of technology. There were resources that could be leveraged, but the strategy had to work within existing constraints.

MakeCode was chosen because of its ease of programming using blocks and the ability to work without a permanent connection. Low-cost electronic components were also used. The teacher's previous preparation in Micro:bit and computational thinking was another element considered. The feasibility of the intervention depended on the relationship between resources, training and available time.

4.5. Organization of the didactic strategy

The strategy was designed using the ASSURE model. This model made it possible to begin with the analysis of the students and the context, to establish objectives, to select methods and means, to use resources, to promote participation and to evaluate the process. The methodological and pedagogical review provided criteria for each of these stages.

Three blocks were organized. The first addressed electronic circuits and components. The second concentrated on programming with Micro:bit. The third integrated the learnings into robotics activities, including the construction of a controlled vehicle, a line-following robot and an obstacle avoidant robot. The sequence was planned from the basic to the integrated, because the students did not have extensive experience in electronics or programming.

The collaborative work was organized through groups and roles. This prevented a single student from concentrating all the tasks and allowed for the distribution of construction, connection, scheduling, and verification. The practical activity was important, but it was related to physics concepts and was not presented as simple technological entertainment.

Methodologically, the description of the intervention was necessary to interpret the results. Two investigations can use Micro:bit and yet apply different treatments. For this reason, the sequence, the contents, the organization of the groups and the products produced were detailed.

4.6. General Instrument Selection

The review showed that the questionnaire was the most frequent instrument in the articles analyzed. However, the frequency of use was not the main reason for selecting the instruments. The correspondence between each questionnaire and the variable to be measured was examined.

For attitude and beliefs towards physics, CLASS was selected. An adaptation of the SMQ-II to the domain of physics, called the Physics Motivation Questionnaire, was used for motivation. Academic performance was measured with 20 physics questions from ICFES public booklets. In this way, each variable had a different instrument.

The application was carried out under standardized conditions. All three instruments were administered in a planned 120-minute session: 20 minutes for CLASS, 40 minutes for PMQ, and 60 minutes for the performance test. Digital forms with mandatory questions and identification codes were used. The use of codes made it possible to relate pre-test and post-test without placing the names of the students in the analysis base.

4.7. Statistical plan

The analysis plan was established after defining the variables, instruments and comparisons. To study the relationship between attitude and performance, and between motivation and performance, Spearman's coefficient was used. This choice was consistent with the ordinal nature of the scales and with the decision not to depend on the assumption of normality.

In the comparison of post-test academic performance between groups, the t-test with Welch correction was used, due to the inequality of variances. For attitude and motivation, individual differences were calculated between post-test and pre-test. These changes were compared between the experimental group and the Mann-Whitney U control group. The significance level was set at $\alpha = .05$.

The same test was not used for all variables. The type of data, the form of comparison and the assumptions determined each procedure. This decision was important because the quasi-experimental design does not require a single analysis; requires selecting the test that corresponds to each question.

4.8. Ethical considerations

The research was carried out with underage students and within an educational institution. For this reason, institutional authorization was requested, voluntary participation was established, informed consent was used, and confidentiality was protected through codes. The project was submitted to the corresponding ethics committee.

Ethical decisions influenced the methodology. It was not possible to force a student to participate or reorganize the groups just to favor experimental control. The protection of participants should be maintained during the collection, storage and presentation of results.

5. Selection and substantiation of the instruments

5.1. Colorado Learning Attitudes about Science Survey

CLASS was developed to measure students' attitudes and beliefs about physics and the way it is learned. The original version contains 42 statements answered on a five-choice Likert scale. The answers are compared with positions considered expert and can be analyzed globally and by categories.

The study by Adams et al. (2006) is the main source for the development of the instrument. Validation included expert review, interviews with students, reliability studies, and statistical analyses with more than 5,000 participants. A previous work by Adams et al. (2005) already described interviews and factor analyses used during the first versions.

A strength of CLASS was that its items were written to be understandable in different courses and to avoid overly specific references to a particular subject. The application was also intended to take little time. These characteristics were appropriate for the thesis population, because the questionnaire had to be applied together with two other instruments.

Subsequent literature showed results that were not identical. Bayar and Karamustafaoglu (2015) worked with 400 ninth graders in Turkey and reported an eight-factor structure, internal consistency between .72 and .84, and test-retest reliability between .85 and .93. Kaltakci-Gurel (2021), with 1,391 students, developed a Turkish adaptation that included experts, interviews, exploratory and confirmatory factor analysis, and proposed a 20-item, three-factor version. Kontro and Buschhüter (2020) analyzed data from 642 Finnish students and obtained a 14-item, three-factor solution. Badeo and Duque (2023) proposed a 29-item, four-factor version with 227 high school students.

The existence of different structures did not lead to the conclusion that CLASS was useless. He showed that their behavior depends on the population and that it must be explained which version is used. In the thesis, the 42-item version and the categories of extended use were preserved, because they facilitated comparison with other studies and corresponded to the operationalization adopted.

Antecedents in Spanish were also considered. Alarcon and De la Garza (2010) used a translation reviewed by bilingual speakers in Mexico. Other Spanish-speaking studies applied CLASS to analyze differences between groups and changes produced by teaching methodologies. This background supported linguistic comprehension and use in close contexts, although it did not replace local assessment.

In the pretest of the thesis, Cronbach's alpha was calculated for the 42 items and $\alpha = 0.7256$ was obtained. This value was considered acceptable for the overall score used. It was interpreted as evidence of internal consistency in the sample and not as a complete revalidation of the instrument.

5.2. Physics Motivation Questionnaire y SMQ-II

The motivation questionnaire used is derived from the Science Motivation Questionnaire. The initial version had 30 items. Glynn et al. (2009) applied it to 770 students of non-scientific careers and found five dimensions: intrinsic motivation and personal relevance, self-efficacy and anxiety, self-determination, career

motivation, and motivation for grades. The overall alpha was .91, but the dimension related to grades presented a lower coefficient, .55.

These difficulties led to a revision of the instrument. Glynn et al. (2011) developed the SMQ-II, composed of 25 items distributed in five five-item scales: intrinsic motivation, career motivation, self-determination, self-efficacy and motivation for grades. The study included students from scientific and non-scientific careers and used exploratory and confirmatory factor analysis.

In the exploratory sample, the five factors explained 67.64% of the variance. The alpha coefficients were .89 for intrinsic motivation, .92 for career motivation, .88 for self-determination, .83 for self-efficacy and .81 for motivation by grades. The global alpha was .92. The review improved the separation between components that were previously linked, especially self-efficacy and anxiety.

The SMQ-II was designed to allow for adaptations by discipline. Versions of this possibility emerged for physics and chemistry. Ersoy and Alička (2016) adapted the instrument as the Physics Motivation Questionnaire with 110 high school students and obtained an overall alpha of .805. Other studies used Rasch models or cross-cultural procedures. You et al. (2018) found good overall functioning, although they pointed out problems in some items. Toma et al. (2023), with 646 Brazilian high school students, obtained a structure of 24 items and four factors, without a clearly differentiated dimension of self-determination.

These variations confirmed the need to review the evidence in each context. In the thesis, a 25-item version adapted to physics and organized in five dimensions was used. The instrument adjusted to the problem because it allowed differentiating intrinsic interest, career projection, autonomous effort, confidence to learn and motivation associated with grades.

The internal consistency calculated with the pretest was $\alpha = 0.8709$. This result supported the use of the score and dimensions in the sample studied. The selection of the PMQ was not justified only by its use in previous research, but also by the documented revision trajectory of the original SMQ, the factor analysis of the SMQ-II and the specific adaptations to physics learning.

5.3. Academic Performance Test

Academic performance was measured using 20 physics questions selected from ICFES public booklets. The items were kept without modifying their wording and were classified into three competencies: inquiry, explanation of phenomena and comprehensive use of scientific knowledge.

The origin of the items offered several advantages. These were questions constructed for a national assessment, with a format known to Colombian students and directly related to the natural sciences competencies evaluated in Saber 11. In addition, they made it possible to measure more than the repetition of formulas, because they included interpretation, explanation and application.

However, bringing together 20 questions from different booklets generated a new test in terms of application. The quality of each original item did not guarantee that the set had high internal consistency in the sample. For this reason, KR-20 was calculated with the results of the pretest and 0.437 was obtained.

The coefficient was low. The thesis interpreted it considering that the test was diagnostic and that some contents had not yet been formally worked on. Questions were not eliminated only to increase the coefficient, because it was intended to preserve the coverage of competencies. Even so, the result forces us to interpret the total score with caution. This situation shows that the standardized origin of the items does not replace the evaluation of the scores obtained in a particular application.

Table 2. Instruments and selection criteria

Variable	Instrument used	Evidence reviewed	Reliability in the thesis	Main reason for selection
Attitude and beliefs towards physics	CLASS, 42 items	Development with experts, interviews, statistical analysis and validations in different countries	$\alpha = 0.7256$	It specifically measures beliefs about physics and its learning and allows comparison with other studies
Motivation towards physics	PMQ derived from SMQ-II, 25 items	AFE, AFC, disciplinary adaptations, Rasch studies and cross-cultural validations	$\alpha = 0.8709$	Differentiate five motivational components related to the research problem
Academic performance in physics	ICFES 20-item test	Items from standardized booklets organized by natural science competencies	KR-20 = 0.437	Curricular correspondence, Saber 11 format and measurement of competencies; cautious interpretation for low certainty

6. Map of the relationship between the review and the final decisions

The methodology can be summarized as a sequence of related decisions. Table 3 shows the main result of this reconstruction. It does not intend to reduce the entire process to a matrix, but to make visible what information was taken into account in each component.

Table 3. Relationship between methodological evidence and decision taken

Component	Information obtained in the review	Criteria applied	Final decision	Sense of decision
Paradigm	The variables could be measured and compared	Correspondence with the question and hypotheses	Positivist	It allowed working with indicators and statistical analysis
Method	A predominance of mixed studies, but the existence of comparative quantitative research	Need to measure changes and relationships	Quantitative	Responded to comparison and association objectives
Design	Little presence of pretest-posttest with control group	Baseline and Comparison Group Need	Quasi-experimental with pretest, post-test, experimental group and control	Adjusted to intact groups and allowed for comparison of changes
Scope	Attitude and motivation were linked to learning	Interest in learning about partnerships in	Correlational and comparative	It allowed the analysis of attitude-performance and

Component	Information obtained in the review	Criteria applied	Final decision	Sense of decision
		addition to the effect of the strategy		motivation-performance
Population and sample	School experiences are carried out with groups already formed	Institutional viability and real participation	Eleventh grade students; $n = 56$; Groups 11A and 11B	He preserved the school organization and reduced arbitrariness by drawing lots of groups
Intervention	Projects, construction, collaboration, and practice were frequent	Consistency with diagnosis and physics content	ASSURE, constructionism, projects, collaborative work and practice	He defined robotics as a strategy and not as an isolated use of a device
Technology	Existence of commercial, open, and low-cost options	Availability, connectivity and teacher training	Micro:bit, offline MakeCode and accessible components	He made the intervention in the rural institution viable
Attitude	CLASS had a history of validation and variations between populations	Disciplinary specificity and local assessment	CLASS de 42 ítems	Comparability and acceptable local α
Motivation	SMQ-II presented psychometric improvement and adaptations to physics	Multidimensionality and accumulated evidence	25-item PMQ	Disciplinary relevance and good local internal consistency
Performance	Need for a curricular and objective measure	Alignment with national competitions	20 ICFES questions	Known format; Low KR-20 reported as limitation
Statistics	Likert data, scores, and comparison of groups and moments	Correspondence between variables, assumptions and contrast	Spearman, Welch y Mann-Whitney	He avoided applying the same test to results of a different nature
Ethics	Participation of minors within a school	Voluntariness, protection and non-unnecessary alteration of groups	Consent, authorization, anonymization, and ethical review	It protected the participants and gave legitimacy to the procedure

The table shows that no decision was justified solely because it was frequent. In some cases, a widespread practice, such as project work, was adopted. In others, a less common alternative was chosen, such as the design with pretest, post-test and control. There were also decisions determined mainly by the reality of the institution, such as the offline use of MakeCode.

7. Discussion

The reconstruction carried out allows us to affirm that the methodological review had a broader function than the elaboration of antecedents. It served to learn about the variety of designs used in educational robotics and, above all, to compare these designs with what was intended to be answered in the thesis. This step avoided selecting a methodology just because it was known or frequent.

Mixed studies were the most numerous in the corpus. This could have led to the adoption of a mixed method. However, the doctoral question focused on evaluating changes in three defined variables and comparing two groups. The quantitative method was better suited to this purpose. The choice did not represent a negative assessment of qualitative methods; it simply responded to the type of evidence required.

Something similar happened with techniques. Interview and observation were frequent, and their usefulness is clear for understanding student interaction, identity in teams, and the experience of participation. However, the research needed measurements that could be applied before and after and that produced comparable scores. For this reason, questionnaires and a performance test were used.

The review also showed that educational robotic expression can mask very different interventions. An experience can last one session or one semester; it can focus on programming, construction, competition, science, or mathematics; it can be extracurricular or part of a subject. This diversity explains why it is not enough to say that research uses robotics. The treatment, sequence, and role of the participants should be described.

In the thesis, we tried to solve this difficulty using the ASSURE model and a sequence organized in blocks. The intervention integrated circuits, programming and construction of robots. This description allows us to understand what the experimental group received and avoids attributing the results to the mere presence of a Micro:bit card.

The rural context had a real methodological weight. The literature showed that access and implementation depend on cost, connectivity, teacher training, and institutional support. The institution had resources, but did not have a stable connection. In this case, the methodology had to take advantage of what was available without designing a proposal dependent on the internet. The technology selection was a contextual decision and not a comparison of brands.

Psychometric review was another important component. CLASS and SMQ-II have strong studies, but adaptations have produced distinct structures. This behavior is to be expected when the age, culture, language, and type of student change. The lesson is not to reject instruments, but to avoid an automatic transfer of their properties.

The thesis calculated internal consistency with its own data. CLASS obtained an acceptable value and PMQ a good value. The performance test showed a low KR-20. This last result should not be hidden or explained solely by the origin of the ICFES questions. Reliability corresponds to the scores obtained by a population in a specific application. Therefore, the coefficient limits the accuracy of the score and should be considered when interpreting results.

The use of artificial intelligence and text mining during the review allowed for the organization of a large set of documents. At the same time, this practice requires verification. A tool can confuse authors, misclassify a design, or attribute a coefficient to a different instrument. In this process, it was maintained as a rule to return to the original document when a statement had a direct impact on the methodology.

The review can serve several functions. It allows us to know alternatives, recognize the conditions under which they have been used and observe problems that other authors found. It also helps justify why an option was discarded. In this case, research based solely on perception would have been insufficient for the impact question; an experiment with individual randomization was not feasible within the school; and a platform that depended on permanent connectivity was not suitable for the institution.

The main utility of the decision map is that it forces us to explain these relationships. Instead of presenting paradigm, design, sample and instruments as separate sections, it shows how each choice depends on the previous one. If you modify the question, you probably change the layout; If you change the way you measure the variable, you can also change the analysis.

8. Procedure derived from the case

Based on the experience described, a ten-step route is proposed. It is not presented as a validated model for any research, but as a way of organizing the work of review and design.

1. **Specify what you want to know.** Before looking for a method, it must be defined whether the research aims to describe, understand, relate, predict or compare. The verb of the objective guides the type of evidence needed.

2. **Look for thematically close studios.** The review should include research on the same phenomenon, but also methodologically comparable works, even if they use other school content.

3. **Record the methodology of each study.** It is convenient to organize population, design, variables, instruments, procedure and analysis in a matrix. This makes it easier to compare and avoids relying on memory.

4. **Recognize trends without turning them into rules.** A frequent method may be helpful, but its popularity doesn't prove that it's appropriate for the new question.

5. **Relate the options to the context.** Time, resources, connectivity, researcher training,

institutional organization and characteristics of the participants must be considered.

6. **Define the variables and their way of measurement.** An instrument cannot be selected before clarifying which aspect of the construct is to be measured and how the score will be interpreted.

7. **Review the evidence of the instruments separately.** Content validity, internal structure, consistency, stability, and linguistic adaptation must be differentiated. It is also necessary to evaluate the instrument with one's own sample when possible.

8. **Build the relationship between question, design and analysis.** Comparison, correlation and estimation of changes require different procedures. Analysis should be planned before data is collected.

9. **Check viability and ethics.** A methodology may be technically appropriate and may not be possible or acceptable in the school context. Participant protection is part of the design.

10. **Record the reasons for each decision.** The researcher must be able to explain which alternatives he examined, why he chose one, and what limitations he retains. This register facilitates the writing and evaluation of the study.

9. Limitations

The first limitation corresponds to the original search. Scopus was used, a time window from 2018 to 2023 and a selection that took into account citations. This procedure facilitated the management of the corpus, but it could leave out recent publications, regional documents and studies not indexed in that database.

The second relates to classification. Some articles combine several designs, techniques or pedagogical fundamentals. Grouping them as qualitative, quantitative, or mixed simplifies characteristics that can be more complex. For this reason, the percentages should be interpreted as corpus trends and not as results of a meta-analysis.

The third limitation corresponds to the use of artificial intelligence in the document

organization phase. These tools speed up extraction, but they can cause errors. Their use requires manual revision and contrast with fonts, especially when it comes to psychometric values or design features.

The fourth is psychometric. The CLASS and SMQ-II studies show that the factor structure can vary between populations. Local internal consistency provided information, but it does not replace a complete validation in rural Colombian students. In addition, the low KR-20 of the performance test reduces the accuracy of the score and calls for caution.

The fifth limitation pertains to doctoral research. The sample consisted of 56 students from a single institution and the groups were initially selected for convenience. The drawing of lots for the function of the groups improved the procedure, but it does not allow the results to be statistically generalized to all rural students.

Finally, additional documents incorporated after the original revision were used for contrast. It should not be said that they guided decisions that had already been made. Maintaining this difference avoids reconstructing the process as if all literature had been available from the beginning.

10. Conclusions

The methodological review allowed the methodology of the thesis to be built gradually. The articles were not used solely to claim that educational robotics produces benefits. We also reviewed how the studies had been conducted, what variables they measured, what instruments they used, and what conditions accompanied the interventions.

The quantitative method and quasi-experimental design were selected for their relationship to the

research question and not because they were the predominant options. The pre-test-posttest measurement and the control group made it possible to compare changes within the real possibilities of an institution that worked with already trained courses.

The review also helped separate variables that could be confounded. Attitude, motivation and performance were measured with different instruments. CLASS and PMQ were selected after examining their development and adaptations, and internal consistency in the sample was calculated. The ICFES test provided curricular relevance, although its low KR-20 was retained as an important limitation.

The rural context influenced the selection of participants, the technology and the procedure. The availability of Micro:bit and computers, along with limited connectivity, led to a strategy that could be developed offline. In this way, the review provided options, but the final decision depended on the institutional conditions.

Generally speaking, methodological review means asking not only what other researchers found, but how they produced those results and whether their decisions are relevant to a new study. The frequency of a method serves as a reference, but it does not replace the analysis of the question, variables, population, resources, and ethics.

The main contribution of this article is to show the chain that united the literature with the final methodology. Making this chain explicit improves consistency and allows the reader to know why each component was chosen. Methodological review can therefore be used as a tool to design, justify and evaluate educational research.

References

- Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics - Physics Education Research*, 2(1), 010101. <https://doi.org/10.1103/PhysRevSTPER.2.010101>

- Akman Selcuk, N., Kucuk, S., & Sisman, B. (2024). Does really educational robotics improve secondary school students' course motivation, achievement and attitude? *Education and Information Technologies*, 29, 23753–23780. <https://doi.org/10.1007/s1586-603-1586-603>
- Alarcon, H., & De la Garza, J. (2010). Assessing students' attitudes in a college physics course in Mexico. *AIP Conference Proceedings*, 1289, 129–132. <https://doi.org/10.1063/1.3515178>
- Anwar, S., Bascou, N. A., Menekse, M., & Kardgar, A. (2019). A systematic review of studies on educational robotics. *Journal of Pre-College Engineering Education Research*, 9(2), 19–42. <https://doi.org/10.7771/586-603.1223>
- Badeo, J. M. O., & Duque, D. A. G. (2023). Reliability and validity of Colorado Learning Attitudes about Science Survey (CLASS). *International Journal of Research in STEM Education*, 5(1), 95–105. <https://doi.org/10.31098/ijrse.v5i1.1191>
- Bayar, A., & Karamustafaoğlu, O. (2015). The Colorado Learning Attitudes about Science Survey (CLASS): The study of validity and reliability. *International Journal of Assessment Tools in Education*, 2(1), 40–57. <https://doi.org/10.21449/ijate.239564>
- Bredder, E., Wilkens, K., Espinal, A., Vieira, C., Harris, P., Bailey, J., & Chiu, J. (2026). Roversa: A low-cost, multilingual robotics platform for K-12 computer science integration. *Educational Technology Research and Development*. <https://doi.org/10.1007/s1586-603-1586-603>
- Camargo, C., Gonçalves, J., Conde, M. Á., Rodríguez-Sedano, F. J., Costa, P., & García-Peñalvo, F. J. (2021). Systematic literature review of realistic simulators applied in educational robotics context. *Sensors*, 21(12), 4031. <https://doi.org/10.3390/s21124031>
- Chang, C. C., & Chen, Y. (2022). Using mastery learning theory to develop task-centered hands-on STEM learning of Arduino-based educational robotics: Psychomotor performance and perception by a convergent parallel mixed method. *Interactive Learning Environments*, 30(9), 1677–1692. <https://doi.org/10.1080/10494820.2020.1741400>
- Chatzopoulos, A., Kalogiannakis, M., Papadakis, S., & Papoutsidakis, M. (2022). A novel, modular robot for educational robotics developed using action research evaluated on Technology Acceptance Model. *Education Sciences*, 12(4), 274. <https://doi.org/10.3390/educsci12040274>
- Costelha, H., & Neves, C. (2018). Technical database on robotics-based educational platforms for K-12 students. In 2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC) (pp. 167–172). IEEE. <https://doi.org/10.1109/ICARSC.2018.8374178>
- Dietz, G., Chen, J. K., Beason, J., Tarrow, M., Hilliard, A., & Shapiro, R. B. (2022). ARtonomous: Introducing middle school students to reinforcement learning through virtual robotics. In *Proceedings of Interaction Design and Children* (pp. 430–441). ACM. <https://doi.org/10.1145/3501712.3529736>
- Dorotea, N., Piedade, J., & Pedro, A. (2021). Mapping K-12 computer science teacher's interest, self-confidence, and knowledge about the use of educational robotics to teach. *Education Sciences*, 11(8), 443. <https://doi.org/10.3390/educsci11080443>
- Ersoy, A. F., & Aliçka, Y. (2016). Motivation to learn in secondary school physics. *Beder Journal of Educational Sciences*, 12, 1–10.
- Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science Motivation Questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, 48(10), 1159–1176. <https://doi.org/10.1002/tea.20442>
- Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science Motivation Questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching*, 46(2), 127–146. <https://doi.org/10.1002/tea.20267>
- González-Fernández, M. O., Flores-González, Y. A., & Muñoz-López, C. (2021). Overview of educational robotics in favor of STEAM learning. *Eureka Journal on Science Teaching and Dissemination*, 18(2), 2301. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2021.v18.i2.2301
- Hennessy Elliott, C. (2020). “Run it through me”: Positioning, power, and learning on a high school robotics team. *Journal of the Learning Sciences*, 29(4–5), 598–641. <https://doi.org/10.1080/10508406.2020.1770763>

- Ioannou, A., & Makridou, E. (2018). Exploring the potentials of educational robotics in the development of computational thinking: A summary of current research and practical proposal for future work. *Education and Information Technologies*, 23(6), 2531–2544. <https://doi.org/10.1007/s1586-603-9729-z>
- Ioannou, A., Socratous, C., & Nikolaedou, E. (2018). Expanding the curricular space with educational robotics: A creative course on road safety. In *Lecture Notes in Computer Science* (Vol. 11082, pp. 537–547). Springer. https://doi.org/10.1007/586-603-586-603586-603_42
- Kaltakci-Gurel, D. (2021). Turkish adaptation and psychometric evaluation of the Colorado Learning Attitudes About Science Survey (CLASS) in physics. *The European Educational Researcher*, 4(3), 355–372. <https://doi.org/10.31757/euer.435>
- Kert, S. B., Erkoç, M. F., & Yeni, S. (2020). The effect of robotics on sixth graders' academic achievement, computational thinking skills and conceptual knowledge levels. *Thinking Skills and Creativity*, 38, 100714. <https://doi.org/10.1016/j.tsc.2020.100714>
- Kontro, I., & Buschhüter, D. (2020). Validity of Colorado Learning Attitudes about Science Survey for a high-achieving, Finnish population. *Physical Review Physics Education Research*, 16(2), 020104. <https://doi.org/10.1103/PhysRevPhysEducRes.16.020104>
- Kucuk, S., & Sisman, B. (2018). Pre-service teachers' experiences in learning robotics design and programming. *Informatics in Education*, 17(2), 301–320. <https://doi.org/10.15388/infedu.2018.16>
- Lopez-Caudana, E., Ramirez-Montoya, M. S., Martínez-Pérez, S., & Rodríguez-Abitia, G. (2020). Using robotics to enhance active learning in mathematics: A multi-scenario study. *Mathematics*, 8(12), 2163. <https://doi.org/10.3390/math8122163>
- Luo, W., Wei, H. R., Ritzhaupt, A. D., Huggins-Manley, A. C., & Gardner-McCune, C. (2019). Using the S-STEM Survey to evaluate a middle school robotics learning environment: Validity evidence in a different context. *Journal of Science Education and Technology*, 28(4), 429–443. <https://doi.org/10.1007/s1586-603-09773-z>
- Manikutty, G., Chilakapati, U., Frey, L. M., Vennila, V., Natarajan, A., & Rao, R. B. (2019). I can also make robots! Inspiring rural Indian children to learn robotics. In 2019 IEEE Tenth International Conference on Technology for Education (T4E) (pp. 78–85). IEEE. <https://doi.org/10.1109/T4E.2019.00022>
- Meador, A., Raygoza, A., Irvin, M., Starrett, A., Quiroz, B., & Cartiff, B. (2025). A cross-case analysis of rural robotics teacher leaders' identity development. *Frontiers in Education*, 10, 1578584. <https://doi.org/10.3389/feduc.2025.1578584>
- Pedersen, B., Larsen, J. C., & Nielsen, J. (2020). The effect of commercially available educational robotics: A systematic review. In *Advances in Intelligent Systems and Computing* (Vol. 1023, pp. 14–27). Springer. https://doi.org/10.1007/586-603-586-603586-603_2
- Pivetti, M., & Dusi, V. (2025). Measuring STEM engagement among middle school children involved in educational robotics: A narrative review. *Discover Education*, 4, 253. <https://doi.org/10.1007/s4586-603-00547-x>
- Sisman, B., & Kucuk, S. (2019). An educational robotics course: Examination of educational potentials and pre-service teachers' experiences. *International Journal of Research in Education and Science*, 5(2), 510–531.
- Sophokleous, A., Christodoulou, P., Doitsidis, L., & Chatzichristofis, S. A. (2021). Computer vision meets educational robotics. *Electronics*, 10(6), 730. <https://doi.org/10.3390/electronics10060730>
- Sullivan, A., & Bers, M. U. (2019). VEX robotics competitions: Gender differences in student attitudes and experiences. *Journal of Information Technology Education: Research*, 18, 97–112. <https://doi.org/10.28945/4193>
- Toma, R. B., Cordeiro Bizerra, A. M., Yáñez, I., & Meneses Villagrà, J. Á. (2023). Cross-cultural adaptation of the Science Motivation Questionnaire II (SMQ-II) for Portuguese-speaking Brazilian secondary school students. *Revista Latinoamericana de Psicología*, 55, 109–119. <https://doi.org/10.14349/rlp.2023.v55.13>
- You, H. S., Kim, K., Black, K., & Min, K. W. (2018). Assessing science motivation for college students: Validation of the Science Motivation Questionnaire II using the Rasch-Andrich rating scale model. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(4), 1161–1173. <https://doi.org/10.29333/ejmste/81821>
- Yepes, I., Barone, D. A. C., & Porciuncula, C. M. D. (2022). Use of drones as pedagogical technology in STEM disciplines. *Informatics in Education*, 21(1), 201–233. <https://doi.org/10.15388/infedu.2022.08>