Preprints.org

Article

Evaluating the Impact of an Educational Intervention Using Project-Based Learning on Post-Pandemic Recovery in Rural Colombia

Mercedes Carmen Arrieta-Cohen , Luz Angela Torres-Arizal , Ricardo León Gómez-Yepes [*]

Article

Evaluating the Impact of an Educational Intervention Using Project-Based Learning on Post-Pandemic Recovery in Rural Colombia

Mercedes Carmen Arrieta-Cohen [1], Luz Angela Torres-Arizal [1] and Ricardo León Gómez-Yepes [2,*]

[1] Center of Science and Technology of Antioquia (CTA)
[2] Universidad de Antioquia
[*] Correspondence: rleon.gomez@udea.edu.co

Abstract: This study evaluates the impact of a Project-Based Learning (PBL) intervention on post-pandemic educational recovery in rural Colombia, focusing on student competencies in mathematics, language, science, and 21st-century skills. Conducted in rural schools, the intervention aimed to address significant learning gaps exacerbated by the COVID-19 pandemic by providing teacher training and direct student support. A pretest-posttest single-group design was used to assess the effectiveness of the intervention, with standardized tests measuring academic competencies and an analytical rubric evaluating 21st-century skills. The results indicate significant improvements in math, language, and science test scores, with notable gains in problem-solving, collaborative work, communication, and critical thinking. However, a decline in creativity scores highlights the need for a stronger emphasis on fostering creativity within the PBL framework. Gender differences were observed, with female students generally outperforming males, suggesting the need for tailored instructional approaches. The study's limitations, including the absence of a control group, non-random sampling, and the use of subjective assessment methods, are acknowledged, with recommendations for future research to address these issues. Despite these limitations, the findings underscore the potential of PBL to enhance student learning outcomes in rural settings, offering valuable insights for educators and policymakers aiming to support educational recovery and development in similar contexts. Further research is recommended to explore the long-term effects of PBL and to refine the intervention for broader implementation.

Keywords: Project-Based Learning; rural education; 21st-Century Skills; Evidence-Centered Design; Depth of Knowledge; program evaluation; primary education; science skills; math skills; language skills

1. Introduction

In the wake of the Covid-19 pandemic, education in Latin America and the Caribbean has faced a severe crisis, reflecting a loss of more than a decade of progress in learning. Even before the pandemic, over one-third of the region's children and adolescents (35 million) were not reaching the minimum proficiency level in reading, and more than half (50 million) were failing to meet learning standards in mathematics [1]. The pandemic has only exacerbated these challenges, making the need of effective educational interventions more urgent than ever.

The 2022 PISA assessment highlights the ongoing challenges in Colombian students' academic proficiency, showing both progress and areas where improvement is still needed. Although there have been gains since 2006 in mathematics, science, and reading, significant gaps persist when compared to OECD averages. In mathematics, Colombia's average score increased by 13 points, yet it continues to fall substantially below the OECD benchmark. Similarly, while there have been improvements in science, recent trends suggest that progress has stalled, with scores still not reaching

the OECD average. Reading scores have improved but continue to lag behind the standards set by the OECD [2].

These statistics underscore the formidable challenges faced by the educational system. In response to these learning gaps, the [institution name redacted for review], a nonprofit organization based in [city redacted for review], Colombia, with funding from [Funder name redacted for review], developed and implemented an educational intervention called [Intervention name redacted for review]. The intervention ran from 2010 to 2023 and focused on improving educational outcomes in rural areas by offering training to primary school teachers and a series of remedial sessions specifically tailored for third graders.

The teacher training covered the design, implementation, and evaluation of PBL activities aimed at enhancing children's competencies in Spanish language, mathematics, science, and 21st-Century Skills [3], and the remedial activities were designed to elevate students' performance to grade-level standards in Spanish language, math, and science. By focusing on both teacher training and direct student support, the intervention sought to address educational deficiencies holistically, fostering an environment where students could recover lost ground and develop essential academic skills. This dual approach not only equipped teachers with the necessary tools and methodologies to implement effective PBL but also provided immediate, targeted support to students who had fallen behind due to the disruptions caused by the pandemic.

This article focuses on the results from the 2022 cohort of 3rd graders, who were the first group to return to in-person learning following the prolonged school closures caused by the COVID-19 pandemic. The primary purpose of this study is to evaluate the impact of Project-Based Learning (PBL) on student performance and the development of 21st-century skills within this unique post-pandemic educational context.

Assessing the effectiveness of educational interventions is essential for enhancing the basic competencies of children in rural areas and for informing decisions about resource allocation and policy development [4]. This cohort presents a critical opportunity to examine how PBL can address learning losses and facilitate skill development during a period of recovery and adjustment. The analysis aims to contribute to the existing literature on the effectiveness of PBL, particularly in fostering academic resilience and adaptability in the face of unprecedented educational challenges.

2. The Intervention

The intervention [name redacted for review] was designed to improve the pedagogical practices of primary school teachers in rural areas of the [region redacted for review] region of Colombia, with the goal of strengthening student competencies in the core subjects such as mathematics, Spanish, and science. Over the 13 years of the program's implementation, it reached approximately 10,500 students and 2,500 teachers. The intervention focused on training and supporting teachers in designing, implementing, and evaluating PBL strategies in their classrooms. Additionally, the program promoted the development of 21st-century skills to enhance the quality of education in rural schools.

2.1. Pedagogical Foundations of the Intervention

The educational intervention was founded on an active learning approach [5–7] that emphasized the development of both foundational competencies in language, mathematics, and science, as well as essential 21st-century skills. This dual focus aimed to prepare students not only with the academic knowledge needed for success but also with the soft skills critical for navigating the complexities of modern life. Central to the intervention's design and implementation was PBL, a pedagogical strategy that actively engages students in real-world projects. PBL encourages learners to take ownership of their education by fostering critical thinking, problem-solving, and collaboration skills, which are integrated with core academic content [8–10].

Research on PBL has demonstrated its effectiveness in improving foundational skills across subjects such as language [11,12], mathematics [13,14], and science [15,16]. PBL engages students in real-world projects and has been associated with increased student engagement and improved

learning outcomes [17,18]. A key element of PBL is its alignment of assessment strategies with project objectives, allowing students to be evaluated on both academic understanding and their ability to apply these skills in practical contexts [19–21].

Similarly, the integration of 21st-century skills into education has become increasingly recognized as essential. These skills, which include critical thinking, collaboration, and digital literacy, are necessary for students' personal development and their ability to succeed in an ever-changing world [3,22–24]. The incorporation of these competencies into educational frameworks aligns with the goals of PBL, as both approaches emphasize the importance of preparing students to meet the complex challenges and opportunities they will face in various aspects of life. Both PBL and 21st-century skills share a focus on preparing students not only academically but also in developing the practical abilities needed for lifelong learning and problem-solving. [23]. PBL supports the acquisition of 21st-century skills by promoting an active learning environment where students are encouraged to think critically, collaborate effectively, and apply their knowledge in diverse situations [25].

However, integrating 21st-century skills into classrooms presents challenges. There is often a gap between recognizing their importance and fully incorporating them into practice [26,27]. Traditional educational approaches, which tend to compartmentalize subjects like mathematics and science, can hinder interdisciplinary learning and limit real-world application opportunities [28].

In contrast, PBL offers a holistic approach, embedding these skills into thematic learning that fosters both academic and interpersonal growth. It encourages students to engage in collaborative projects that require critical thinking, problem-solving, and effective communication, all essential to 21st-century competencies [27]. Research has shown that PBL not only enhances student motivation and engagement but also improves their ability to work collaboratively and think critically [29]. By linking theory to real-world applications, PBL can bridge the gap between knowledge and practice [30].

2.2. Evaluative Foundations

The evaluation of learning outcomes in the intervention draws upon the Evidence-Centered Design (ECD) framework developed by Mislevy and colleagues [31–34], which provides a systematic and structured approach to educational assessment. This framework has been adapted to meet the unique requirements and users of the program. At the core of the ECD framework are three interconnected models, collectively forming what is referred to as the Conceptual Assessment Framework (CAF): the proficiency model, the task model, and the evidence model (Figure 1) [32].
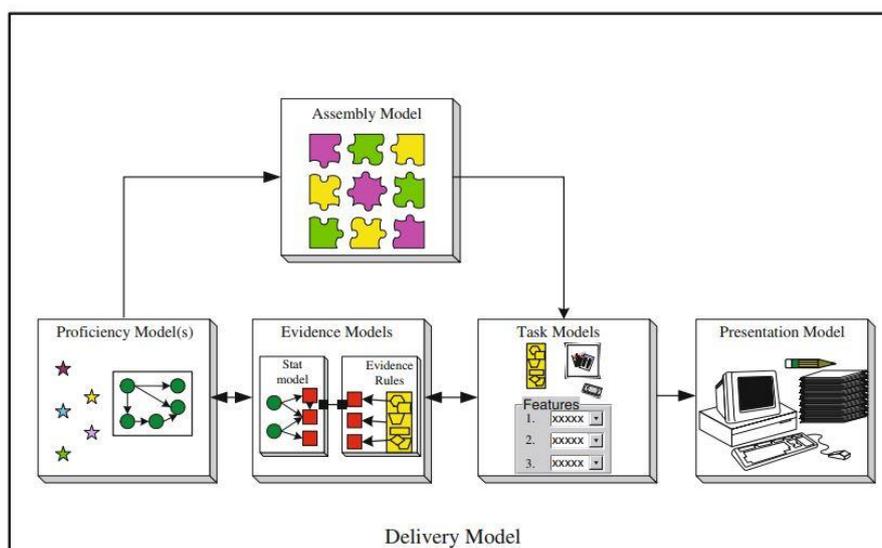

Figure 1. The components of the Evidence-Centered Design [34]. Reprinted with permission from the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The *proficiency model* forms the basis of the assessment strategy, identifying the knowledge, skills, and abilities (KSA) targeted by the intervention. These competencies are aligned with national standards for math, reading, and science education [35]. The proficiency model breaks down these competencies into specific skills and subskills, providing a clear framework for what students need to achieve. This structure guides the design of teacher training and student remediation activities and the subsequent evaluation of student learning.

The *task model* translates latent competencies into observable behaviors by defining the key features of assessment tasks and items. It ensures that these tasks effectively elicit student behaviors that reflect their competencies. In this intervention, the task model followed Norman Webb's Depth of Knowledge (DOK) framework [36–38], which categorizes cognitive demands into various levels and guides task creation. This approach ensures assessments address a range of cognitive processes, including recall, reasoning, application of skills, and deeper conceptual understanding. For example, tasks designed to assess data interpretation from graphs focus on both recalling relevant information and applying interpretation and problem-solving skills.

The *evidence model* connects the latent competencies defined in the proficiency model with the observable behaviors elicited by the task model. It encompasses key elements of measurement, including the development of tests, rubrics, scoring methods, and statistical analyses for establishing students' performance. The tests and rubrics are designed to align with the competencies outlined in the *proficiency model* and the tasks specified by the *task model*. These assessments incorporate multiple item types—such as multiple-choice, short answer, and performance-based tasks—to provide a comprehensive evaluation of student abilities. Clear criteria for evaluating student responses were established to ensure consistency and objectivity by defining specific performance levels. As shown in Figure 2, the intervention assessed students using two instruments: a standardized test measuring competencies and cognitive processes in language, mathematics, and science, and an analytical rubric [39] assessing 21st-century skills. For teachers, an observation rubric was used to track the implementation of PBL-based learning strategies in the classroom.
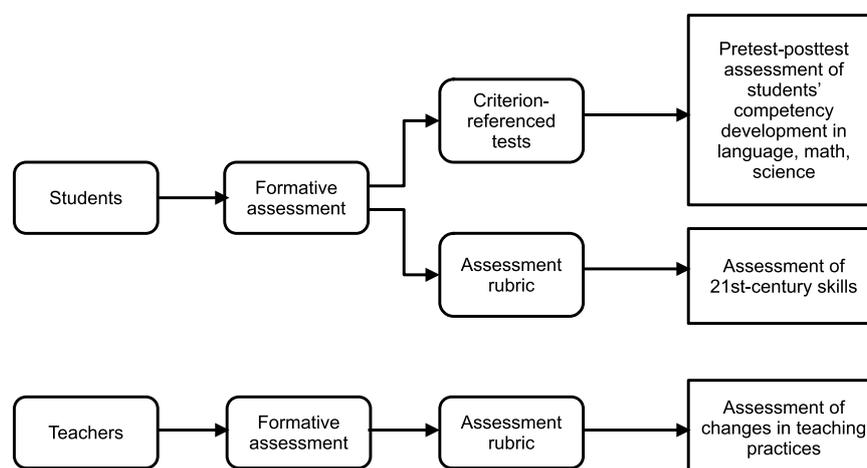
Figure 2. Evaluation framework of student learning and teacher development.

2.2. Implementation of the Intervention

The implementation of the intervention was designed to be comprehensive, involving both teacher development and student support. This approach was organized around three distinct stages.

The program ran from 2010 to 2023, with a new cohort of teachers and students participating each year. This study specifically focuses on the 2022 cohort, offering insights into the program's impact during this period, particularly in the context of post-pandemic educational recovery.

2.2.1. Stage 1: Planning

In the planning phase, participating schools were identified and contacted to establish the foundation for the intervention. The design of the intervention followed a Theory of Change (TOC) perspective [40,41], which mapped out the program's sequence of inputs, activities, outcomes, and outputs, summarized in a logic model shown in Figure 3. This TOC approach helped in identifying key performance indicators and milestones, allowing for effective monitoring and adaptive program management.
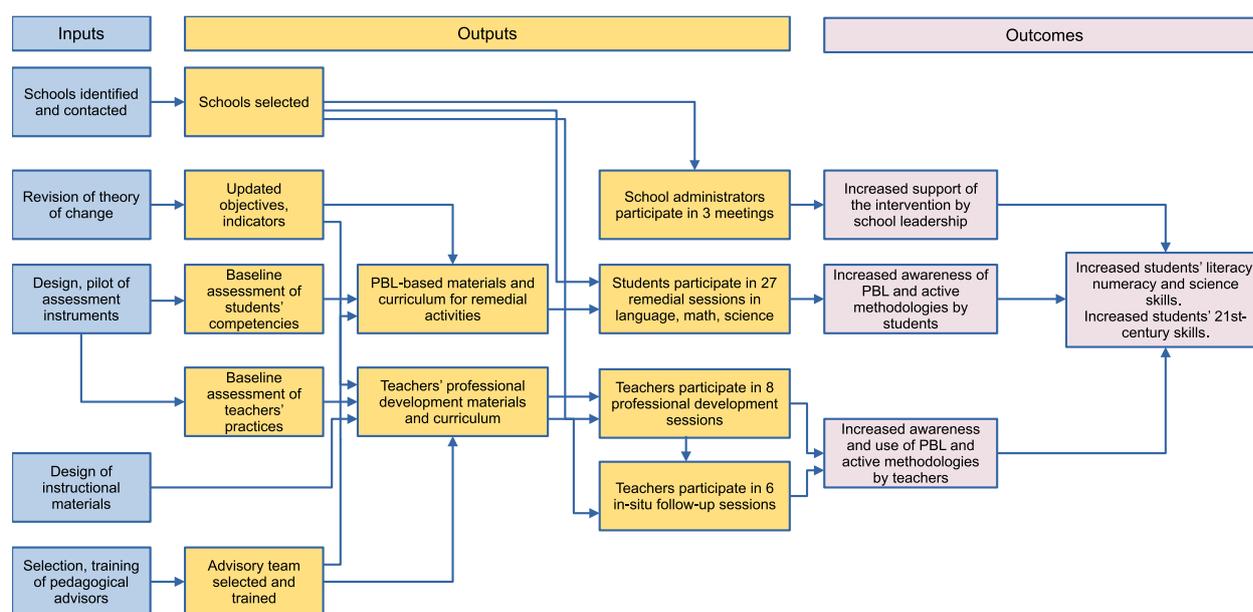

Figure 3. Program's Logic Model.

The TOC was then refined to align with the program's educational goals, and assessment instruments were developed and piloted to measure both student competencies and teacher practices accurately. Instructional materials were created to address the specific needs of the students and teachers participating in the intervention. The selection and training of educational advisors supported the implementation process. Additionally, remediation materials were revised to target specific learning gaps among the students, ensuring the intervention addressed their educational needs.

2.2.2. Stage 2: Baseline Assessment

In the second phase, baseline assessments of students' competencies in math, science, and Spanish language were conducted to establish a starting point for measuring progress. Teachers' instructional and evaluation practices were also evaluated to identify areas needing improvement.

2.2.3. Stage 3: Delivery

The delivery phase involved eight professional development sessions for teachers, focusing on PBL. These sessions covered both theoretical and practical aspects of designing, implementing, and evaluating PBL-based teaching and learning activities within instructional units. Topics included curriculum alignment, project design, student engagement strategies, formative assessment techniques, and integrating 21st-century skills into classroom activities. The training was further

reinforced with six in-situ follow-up sessions to offer continuous support and guidance. School administrators participated in three meetings to ensure alignment and support at the leadership level.

The remedial sessions for students were designed to address specific learning gaps in reading, math, and science, aiming to bring participants up to grade-level standards. These sessions were structured to offer targeted support through individualized instruction, group activities, and problem-solving exercises that reinforced core academic concepts. The curriculum focused on strengthening foundational skills that had been disrupted due to the COVID-19 pandemic, with a particular emphasis on interactive learning and real-world applications.

A total of 27 remedial sessions were conducted, with 24 sessions planned and facilitated by experienced pedagogical advisors hired by the CTA. These sessions employed a combination of direct instruction, hands-on learning activities, and formative assessments to monitor student progress. The remaining three sessions were designed and implemented by the teachers who participated in the professional development program, allowing them to apply their newly acquired skills in PBL. Throughout these sessions, students were encouraged to engage in collaborative tasks, critical thinking, and applied learning to help them build both subject-specific competencies and 21st-century skills. The remedial program aimed not only to improve academic performance but also to foster greater confidence and engagement in the learning process.

Based on the program's theory of change (Figure 3), the inputs and activities were expected to increase literacy, numeracy, and science skills among students. Additionally, teachers were expected to become more aware of and skilled in using PBL and active methodologies in their classrooms, leading to improved instructional and evaluation practices (Figure 2).

At the end of the intervention, students and teachers were assessed to identify the improvements in student competencies and the effectiveness of the professional development for teachers. These assessments aimed to measure the overall impact of the program and to gather insights for future educational initiatives.

3. Materials and Methods

3.1. Participants

As mentioned before, the study focused on the 2022 cohort of third-grade students from the [name or region redacted for review] region of Colombia. The program engaged a diverse group of participants, comprising a total of 287 students, with 150 boys and 137 girls from nine rural schools across seven different municipalities. This distribution ensured a balanced representation of gender and a wide geographic reach, aligning with the program's goal of improving educational outcomes in underserved rural areas. To maintain ethical standards, informed consent was obtained from all participating students' parents or guardians. The names of the schools have been anonymized to protect the privacy of participants and institutions, as shown in Table 1

Table 1. Anonymized Distribution of Program Participants by Location, School, and Gender (N=287).

Municipality	School	N	Percentage (%)
Municipality A	School A	31	10,8
Municipality B	School B	58	20,21
Municipality C	School C	59	20,56
Municipality D	School D	29	10,1
Municipality E	School E	17	5,92
	School F	22	7,67
Municipality F	School G	9	3,14
Municipality G	School H	24	8,36
	School I	38	13,24
	Total	**287**	**100**

3.2. Instruments

A multifaceted assessment strategy was implemented to evaluate program outcomes, measuring students' mastery of core competencies and cognitive abilities in language, mathematics, and science (Table 2). To align assessments with the program's learning objectives, an evidence model was developed based on the competencies outlined in the proficiency model and their corresponding observable behaviors in the task model [32]. This model incorporated a suite of assessments, including standardized pre- and post-tests, a rubric for assessing 21st-century skills, scoring methodologies for each instrument, and statistical analysis to determine the competency level of each student.

Language assessment focused on inferential and critical reading, using a 12-item test to evaluate comprehension, analysis, and critical thinking skills. The science assessment centered on ecological concepts, with a 20-item test assessing students' understanding of abiotic and biotic factors and their interactions. The mathematics assessment measured students' ability to estimate, calculate, and justify their reasoning through 12 questions evaluating computational skills, problem-solving, and logical reasoning (Table 2).

Test items were aligned with Webb's Depth of Knowledge (DOK) framework [36–38], encompassing recall, application, and strategic thinking levels. Recall items tested basic knowledge, application items assessed practical application, and strategic thinking items evaluated the students' capacity to analyze, synthesize, and make decisions based on their understanding. By integrating these components, the Evidence Model created an assessment system that provided a detailed picture of student learning and competencies. This data-driven approach informed instructional improvements and program evaluation.

Table 2. Selected standards addressed by the program [35].

Area	Standard	Evidence	Number of items
Language	Reading Comprehension and Critical Analysis	The student can understand the content of a text by analyzing its structure and employing inferential and critical reading processes.	12
	Text Production and Composition	The student can produce various types of texts (expository, narrative, informative, descriptive, argumentative) while considering grammatical and orthographic conventions.	
Science	Understanding Abiotic and Biotic Interactions	The student can explain the influence of abiotic factors (light, temperature, soil, air) on the development of biotic factors (fauna, flora) within an ecosystem.	20
	Ecological Relationships and Survival	The student can understand and explain the intra- and interspecific relationships between organisms and their environment, emphasizing their importance for survival.	
Mathematics	Estimation and Mathematical Strategies	The student can propose, develop, and justify strategies for making estimates and performing basic operations to solve problems effectively.	12

Data Interpretation and Problem Solving	The student can read and interpret information from frequency tables, bar graphs, and pictograms with scales to formulate and solve questions related to real-world situations.

To complement the quantitative data from standardized tests, the program incorporated a focused assessment of 21st-century skills, prioritizing problem-solving, collaboration, creativity, communication, and critical thinking (Table 3). These competencies, as delineated in the Partnership for 21st Century Skills framework [3], are important for navigating the complexities of contemporary society.

Table 3. 21st Century Skills addressed by the intervention [3].

Skills	Description
Problem Solving	The ability of the student to identify a contextual problem and propose and implement alternative solutions.
Collaboration	The ability of the student to take on established roles, listen to peers, and contribute ideas in an organized manner.
Creativity	The ability of the student to represent the solution to the problem through a product, using various resources for its creation and strategies for its presentation
Communication	The ability of the student to convey ideas or opinions clearly and coherently, both verbally and non-verbally.
Critical Thinking	The ability of the student to analyze, evaluate, and synthesize information in a thoughtful and systematic way. It involves the capacity to question assumptions, identify biases, assess evidence, and consider alternative perspectives.

Classroom observations, guided by a structured analytical rubric [39,42], were used for the evaluation of these skills (Table 4). The rubric offered a framework that transcended traditional academic metrics, allowing for the assessment of skill application in real-world contexts. By operationalizing the assessment of these skills, educators were empowered to evaluate student performance across the five dimensions of 21st-century skills, from problem-solving to effective communication.

Table 4. 21st Century Skills Assessment Rubric.

21st Century Skills	Assessment Criteria	Score (from 0 to 3)
Problem Solving	☐ Identifies a problem ☐ Searches for possible solutions ☐ Implements a solution	
Collaborative Work	☐ Assumes established roles ☐ Listens to peers ☐ Contributes ideas in an organized manner	
Creativity	☐ Represents the solution through a product	

	☐ Utilizes various resources in product creation	
	☐ Applies strategies to present the product	
Communication	☐ Conveys ideas verbally or non-verbally	
	☐ Communicates clear ideas	
	☐ Expresses coherent solutions	
Critical Thinking	☐ Analyzes information	
	☐ Interprets arguments	
	☐ Supports solutions	

Score Definitions: 0 = Beginning, 1 = In Progress, 2 = Achieved, 3 = Extended.

The integration of these assessment components provided a comprehensive perspective on student capabilities, extending beyond cognitive skills to include essential competencies for modern learning. This approach aligns with the program's objectives, emphasizing the development of critical thinking, problem-solving, and collaborative abilities as prerequisites for success in education, the workforce, and life.

3.3. Data Collection Procedures

The intervention employed a pretest-posttest single-group design, incorporating standardized tests and classroom observations. For the language component, N=264 students completed the pretest, N=226 took the posttest, and N=196 had both sets of results for analysis. In science, N=265 students completed the pretest, N=221 took the posttest, with N=190 providing both pretest and posttest results. For mathematics, N=267 students completed the pretest, N=233 took the posttest, and N=201 had both results available for analysis. The tests were administered by CTA advisors in a controlled environment to ensure consistency and reliability. Students were provided with clear instructions, monitored for compliance, and given breaks as needed. The tests were administered in a structured, uniform manner, with strict adherence to time limits.

Additionally, N=268 students were observed at the beginning of the intervention, and N=273 were observed at its conclusion, with N=238 having both pre- and post-intervention observation results. These observations were conducted by CTA advisors who filled out analytical rubrics to assess the application of 21st-century skills during the implementation of PBL activities. The difference in the number of students between the pretest and posttest was due to factors such as absences, changes in enrollment, or logistical challenges that prevented some students from completing both assessments.

Upon completion, tests and rubrics were collected and entered into LimeSurvey [43] for data streamlining, management, formatting, and creation of data dictionaries.

3.4. Data Preparation and Analysis

3.4.1. Cognitive Tests

Student responses were processed to generate a global score for each student in each subject—language, mathematics, and science—using a weighted approach based on the cognitive demands of the assessment items. Each item was categorized as measuring recall, application, or strategic thinking, with weights assigned to reflect the complexity of each item type. After applying these weights, the scores were normalized using z-scores to standardize the distribution of responses.

Subsequently, a global score for each student in each subject was calculated. To facilitate comparison across subjects, these global scores were converted to a T-scale with a mean (M) of 50 and a standard deviation (SD) of 10 [44,45]. This transformation ensured consistency in measuring

student performance across subjects and provided a standardized metric for evaluating individual student achievement within the intervention.

In addition to global scores, a measure of competency level was calculated based on students' performance on each task or item within the tests. The criteria used to determine each performance level are shown in Table 11. These competency levels were established to offer a clearer understanding of students' progression and areas of improvement, enabling the identification of specific strengths and weaknesses, and informing targeted interventions and support.

Changes in global scores between pre- and post-tests were evaluated using paired t-tests for each subject area (reading comprehension, mathematics, and science) to determine whether there were differences in mean scores. To account for potential non-normality in the data, Wilcoxon rank-sum tests were conducted as complementary analyses to the t-tests. These non-parametric tests verified the results obtained from the t-tests and provided additional robustness to the findings. Additionally, a repeated measures ANOVA was implemented to assess overall score changes and explore potential gender-based differences in performance trajectories. This analysis provided insights into the general trends in score changes and identified any significant interactions between gender and measurement conditions. All data analyses and visualizations were conducted using R. [46]

3.4.2. Evaluation of Students' 21st-Century Skills

The 21st-century skills rubric scores were used to generate a global score through a multi-step process. First, raw scores were calculated for each of the five dimensions: problem-solving, collaboration, creativity, communication, and critical thinking. These raw scores were then standardized into z-scores to allow comparability across the different dimensions.

Subsequently, each z-score was converted to a 0-10 scale using a normalization function that adjusted values based on the minimum and maximum scores within each dimension. This ensured that all dimensions were on the same scale, facilitating direct comparisons between them.

Following this, an overall score for 21st-century skills was calculated by summing the standardized scores for each dimension. To enhance interpretability, the overall score was also converted to a 0-10 scale, aligning it with the individual dimension scores. Additionally, students' performance levels for each 21st-century skill dimension were determined by first aggregating individual assessment items into total scores for each competency, and then assigning a performance level based on the total score: "Beginning" for a score of 0, "In Progress" for a score of 1, "Achieved" for a score of 2, and "Extended" for a score of 3.
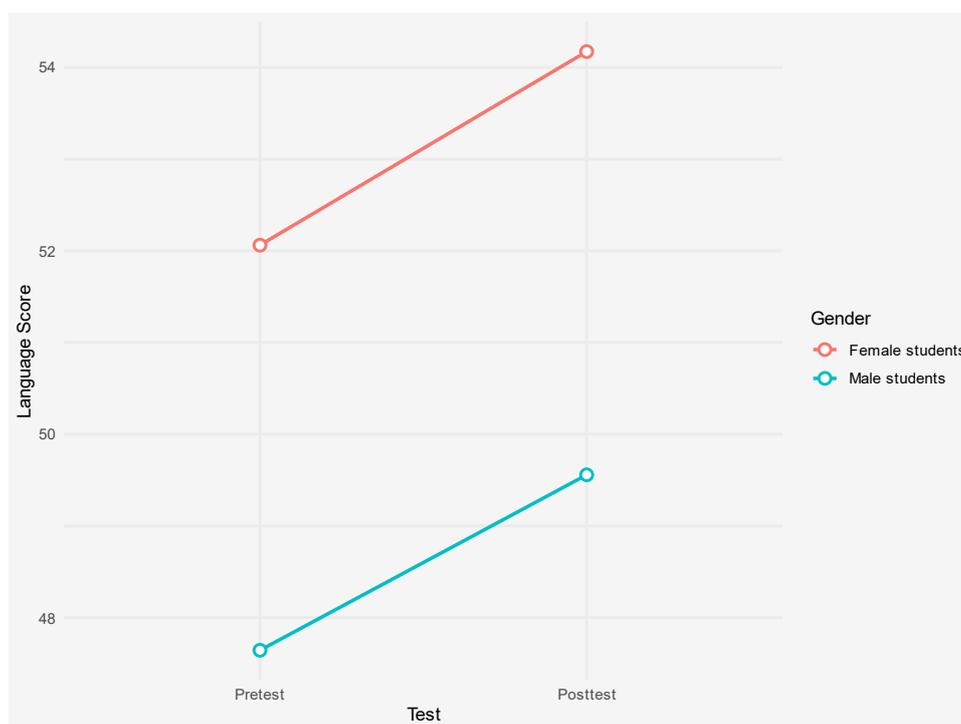
4. Results

4.1. Results of Pretest and Posttest Assessment

This section presents the results of statistical tests conducted to evaluate changes in student performance between pretest and posttest conditions. The analyses included paired t-test and Wilcoxon rank-sum tests to determine significant differences in global test scores, as well as repeated measures ANOVA to examine overall score changes and any gender-based differences in performance.

4.1.1. Language

The assessment results indicate an increase in language test scores from the pretest to the posttest. For girls, the mean score increased by approximately 2 points, rising from 52.1 (SD=10.2) on the pretest to 54.2 (SD=8.9) on the posttest. Similarly, boys' mean scores increased from 47.6 (SD=10.5) on the pretest to 49.6 (SD=9.3) on the posttest (Figure 4).

Figure 4. Change in Language scores from pretest to posttest.

Table 5 below presents the results of the paired samples t-test comparing the language test scores between the pretest and posttest conditions.

Table 5. Paired Samples T-Test Language Test Scores.

Test	Statistic	z	df	p	Effect Size	SE Effect Size
Student	-3.551		195	< .001	-0.254	0.057
Wilcoxon	5440.000	-3.161		0.002	-0.277	0.087

Note. For the Student t-test, effect size is given by Cohen's d . For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

The paired t-test produced a t-statistic of -3.5515 with 195 degrees of freedom and a highly significant p-value (< .001). The effect size, as indicated by Cohen's d, is -0.254 with a standard error of 0.057, suggesting a moderate effect size.

The Wilcoxon test produced a test statistic of 5440.000 with a z-value of -3.161 and a significant p-value (0.002). The effect size for the Wilcoxon test, given by the matched rank biserial correlation, is -0.277 with a standard error of 0.087, also indicating a moderate effect. Overall, the results of both statistical tests for the language test suggest a significant change in test scores between the pretest and posttest conditions.

The results of the analysis of variance (ANOVA) assessing the impact of gender and measurement condition (pretest-posttest) on language test scores are presented in Table 6.

**Table 6.** Impact of Gender and Measurement on Language Scores.

Source of Variation	Df	Sum of Squares	Mean Square	F Value	p-Value
Error: id	1	112	112		

Error: id:measurement	1	1.999	1.999	
Error: Within	386	36851	95.5	
Gender	1	2035	2034.8	21.313	.0000 *
Type of Test	1	396	396.0	4.148	.0424 *
Gender:Type of test	1	1	1.5	0.015	.9014

The analysis revealed statistically significant main effects for both gender and type of measurement condition on language test scores. Specifically, the gender factor demonstrated a significant impact, with an F-value of 21.313 and a p-value of .0000, indicating that gender significantly influenced test scores. Similarly, the type of measurement exhibited a significant main influence, with an F-value of 4.148 and a p-value of 0.0424, signifying its substantial impact on scores. However, the interaction between gender and type of test was not statistically significant (F = 0.015, p = 0.9014), suggesting that the way the scores changed from pretest to posttest did not differ significantly between genders.

4.1.2. Mathematics

The assessment results indicate an increase in math test scores from the pretest to the posttest (Figure 5). For girls, the mean score increased by approximately 2 points, rising from 52.1 (SD=10.2) on the pretest to 54.2 (SD=8.9) on the posttest. Similarly, boys' mean scores increased from 47.6 (SD=10.5) on the pretest to 49.6 (SD=9.3) on the posttest
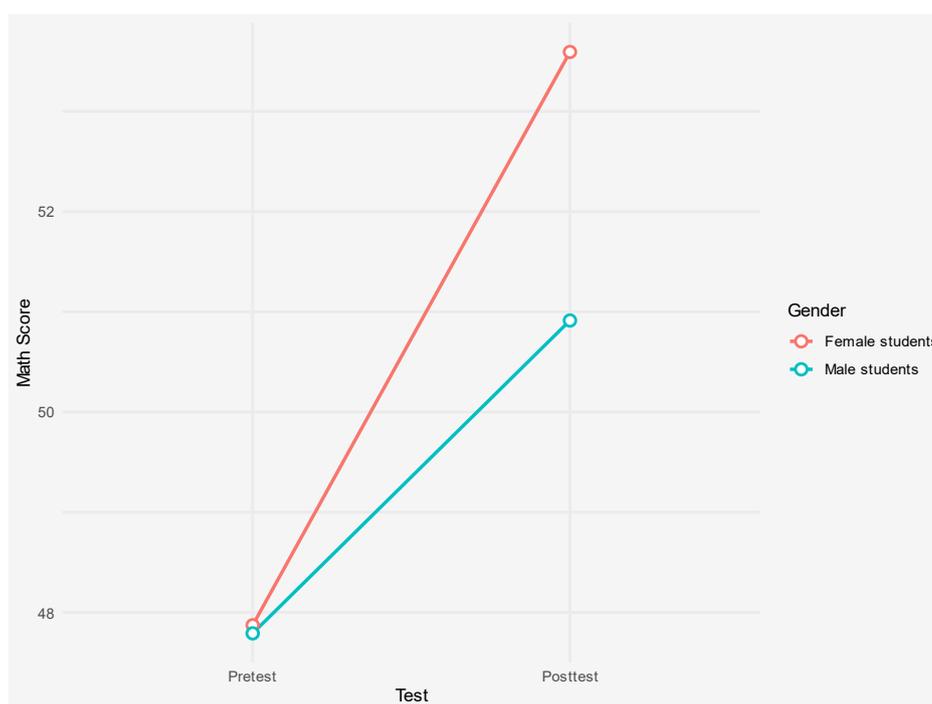
Figure 5. Change in Math Scores from Pretests to Posttest.

As shown in Table 7, The paired t-test yielded a t-statistic of -6.44 with 200 degrees of freedom and a p-value of <.001. This result provides evidence of a significant difference in mean test scores between the two measurement conditions. The Wilcoxon rank sum test yielded a test statistic (W) of 4355 and a p-value of <.001. This result confirms the results to the paired t-test and provides strong evidence of a significant difference in test scores between the two measurement conditions

Table 7. Paired Samples T-Test Math Test Scores.

Test	Statistic	z	df	p	Effect Size	SE Effect Size	95% CI for Effect Size	
							Lower	Upper
Student	-6.44		200	< .001	-0.454	0.068	-0.599	-0.308
Wilcoxon	4355	-5.824		< .001	-0.494	0.085	-0.609	-0.358

Note.   For the Student t-test, effect size is given by Cohen's d . For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

The repeated measures ANOVA test showed that the type of test (pretest or posttest) had a significant impact on math test scores. The F-value of 19.312 and a p-value of less than 0.001 mean that the difference in scores before and after the intervention is statistically significant (Table 8). This result indicates that the intervention had a meaningful effect on the students' math performance.

Table 8. Impact of Gender and Measurement on Math Scores.

Source of Variation	Df	Sum of Squares	Mean Square	F Value	p-Value
Error: Student	1	1.779	1.779		
Error: Student:Measurement	1	49.96	49.96		
Error: Within	396	38003	96.0		
Gender	1	192	191.8	1.999	>.05
Type of test	1	1853	1853.3	19.312	<.001
Gender:Type of Test	1	189	189.4	1.974	>.05

4.1.3. Science

The assessment of scientific skills showed a significant change between pretest and posttest scores (Figure 6).
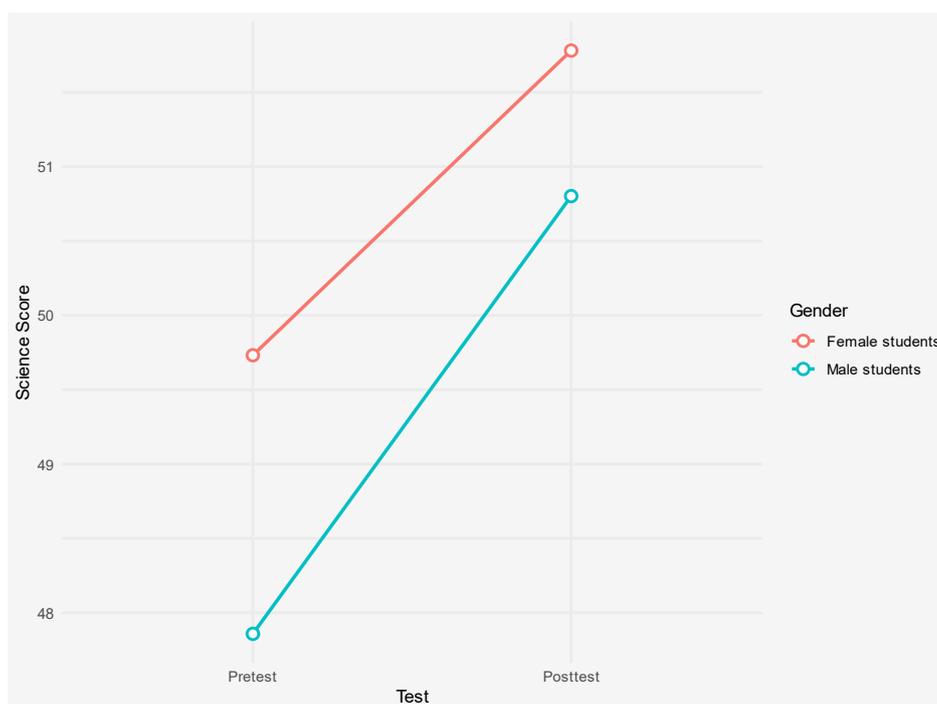
Figure 6. Change in Science Scores from Pretest to Posttest.

The analysis produced a test statistic (W) of -2.903 with 189 degrees of freedom and a p-value of 0.004135. The p-value indicates a significant difference in mean test scores between the two measurement conditions. Both the Wilcoxon rank sum test and the paired t-test provide evidence of a significant change in science test scores between the pretest and posttest conditions. (Table 9).

Table 9. Summary of Test Statistics, Effect Sizes, and Confidence Intervals for Math Test Scores.

							95% CI for Effect Size	
Test	**Statistic**	**z**	**df**	**p**	**Effect Size**	**SE Effect Size**	**Lower**	**Upper**
Student	-2.903		189	0.004	-0.211	0.085	-0.354	-0.067
Wilcoxon	6322	-2.277		0.023	-0.197	0.086	-0.354	-0.03

Note. For the Student t-test, effect size is given by Cohen's *d*. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

The results of the analysis of variance (ANOVA) indicate a significant shift in science test scores between the pretest and posttest conditions, as evidenced by a highly significant F-value ($F(1, 49.96)$ = 17.997, p < 0.001). Specifically, there is a significant variation in scores between the two measurement conditions (Table 10).

For female students, the scores increased from an average of 47.5 in the pretest to 52.7 in the posttest, while male students experienced an increase in scores from 48.2 in the pretest to 51.5 in the posttest. These changes were observed within each gender group.

However, when considering the influence of gender on test scores, the ANOVA results indicate that it did not appear to significantly contribute to the observed variations in test scores ($F(1, 1.892)$ = 0.077, p = 0.781).

The ANOVA results demonstrate a significant change in science test scores between the pretest and posttest conditions, with improvements observed for both F and male students. However, gender itself did not emerge as a statistically significant factor affecting these score differences.

Table 10. Summary of ANOVA results for science scores.

Source of Variation	Df	Sum of Squares	Mean Square	F Value	p-Value
Error: studentid	1	1.892	1.892	0.077	0.781
Error: studentid:medicion	1	49.96	49.96	17.997	2.77e-05 *
Error: Within	390	37800	96.9		
sexo	1	7	7.5	0.077	0.781
medicion	1	1744	1744.3	17.997	2.77e-05 *
sexo:medicion	1	104	103.5	1.068	0.302

Table 11. Established Performance Levels Based on Students' Task Completion.

Performance Level	Probability Threshold	Description
Beginning	Probability of answering recall level questions correctly is less than 50%.	Students at this level are just starting to grasp basic concepts and often struggle with foundational knowledge.
In Progress	Probability of answering recall level questions correctly is greater than 50%, and application-level questions correctly is less than 50%.	Students at this level show some understanding of basic concepts but need further development in applying their knowledge to practical scenarios.
Achieved	Probability of answering application-level questions correctly is greater than 50%, and strategic thinking level questions correctly is less than 50%.	Students at this level have met the standards for their age and grade, demonstrating a good grasp of both basic and applied concepts.
Extended	Probability of answering strategic thinking level questions correctly is greater than 50%.	Students at this level exhibit a strong understanding of both basic and advanced concepts, applying their knowledge effectively and engaging in higher-order thinking tasks.

These statistics provide a detailed view of the test score distribution within different groups. Notably, the mean scores for both genders tend to be slightly higher under the posttest measurement condition compared to the baseline with corresponding standard deviations indicating variability in scores.

4.2. Assessment of Students' Competency Level by Subject

4.2.1. Language Competency Level

The results of the Language component of the intervention reveal changes in student performance levels from the pretest to the posttest (Figure 7). At the baseline (Pretest), the distribution of students was as follows: 27.0% of students were at the Extended level, 22.4% had Achieved the expected competencies, 30.6% were at the Beginning level, and 19.9% were In Progress.
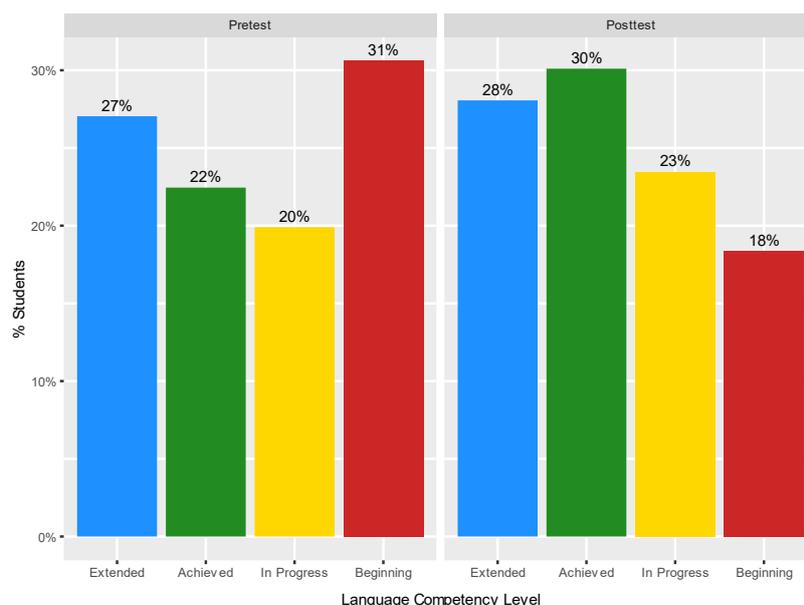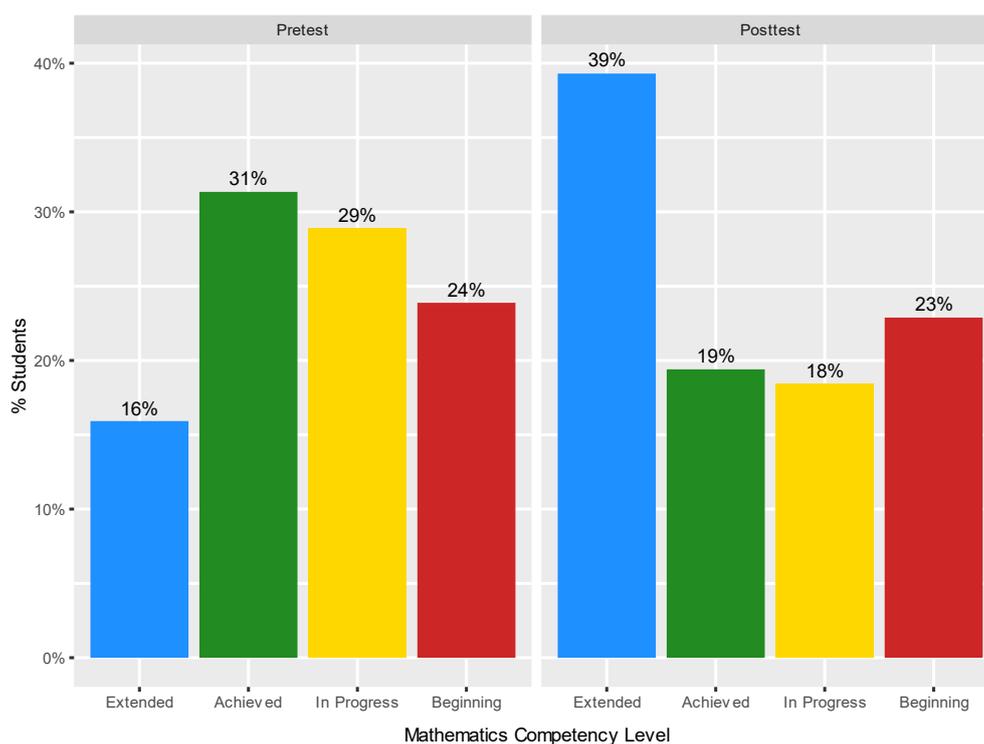
Figure 7. Language Component, General Results by Competency Level (N=196).

By the endline assessment, the distribution showed a shift towards higher performance levels. The percentage of students at the Extended level increased slightly to 28.1%, while those in the Achieved category saw a more significant rise to 30.1%. Meanwhile, the proportion of students at the Beginning level decreased to 18.4%, and those In Progress increased to 23.5%.

These results suggest an overall improvement in language competencies, with more students advancing to the higher performance levels (Achieved and Extended) by the endline assessment. The decrease in the Beginning level indicates that fewer students remained at the lowest level of performance, while the increase in the In Progress level suggests that some students were still transitioning toward achieving proficiency.

4.2.2. Mathematics Competency Level

As shown in Figure 8, prior to the intervention, 31.3% of students had Achieved the expected competencies, 28.9% were In Progress, 23.9% were at the Beginning level, and 15.9% were at the Extended level.

Figure 8. Mathematics Component, General Results by Competency Level (N=201).

By the posttest assessment, the distribution of performance levels showed notable changes. The percentage of students at the Extended level increased significantly to 39.3%, reflecting a considerable improvement. However, the percentage of students who Achieved the expected competencies decreased to 19.4%. Meanwhile, the proportion of students In Progress decreased to 18.4%, and those at the Beginning level also saw a slight reduction to 22.9%.

4.2.3. Science Competency Level

The results of the science component of the intervention, as shown in Figure 9, indicate changes in student performance across the four competency levels. At the baseline, the distribution of students was as follows: 22.6% of students were at the Beginning level, 32.6% were In Progress, 24.2% had Achieved the expected competencies, and 20.5% were at the Extended level. By the endline assessment, the distribution shifted, with 18.9% of students remaining at the Beginning level, a decrease from the baseline. The percentage of students In Progress remained constant at 24.2%, while those in the Achieved category increased to 25.8%. Notably, the percentage of students at the Extended level rose significantly to 31.1%. These results suggest a positive shift in student performance, with a notable increase in the proportion of students reaching the higher performance levels (Achieved and Extended) by the endline assessment. The decrease in the Beginning level indicates that fewer students remained at the lowest level of performance, reflecting overall improvement in the science competencies targeted by the intervention.
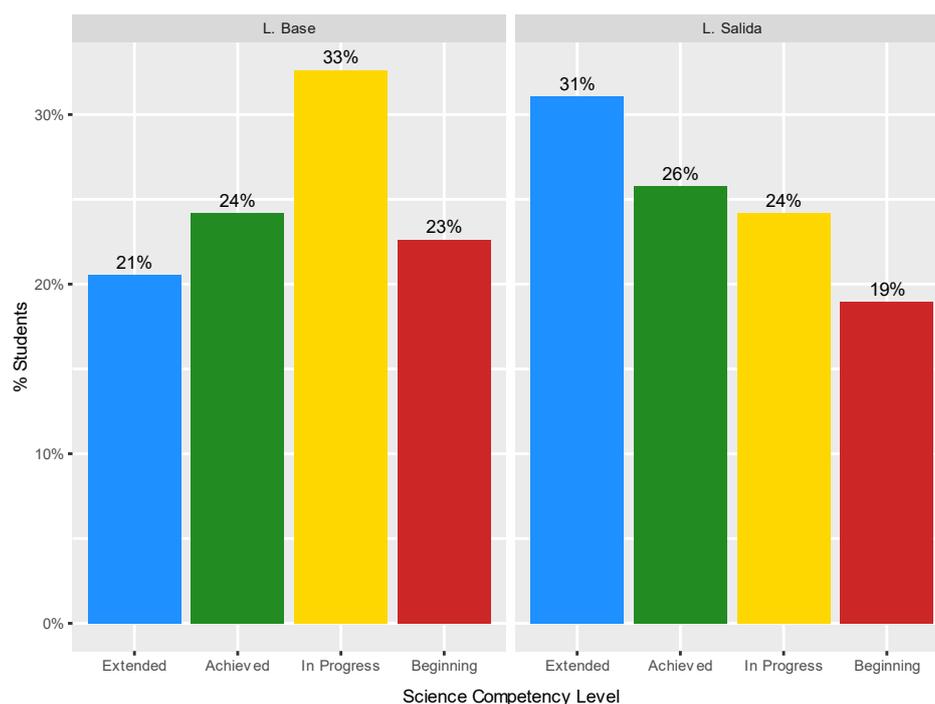
Figure 9. Science component, general results by competency level (N=190).

These results suggest a significant shift towards higher performance levels, particularly with the large increase in students reaching the Extended level by the posttest assessment. The decrease in both the Achieved and In Progress categories may indicate that some students who were previously meeting or nearly meeting the expected standards advanced further to the Extended level, while the slight reduction in the Beginning level suggests an overall improvement in foundational skills.

4.3. Assessment of 21st-Century Skills

As illustrated in Figure 10, the changes in participants' standardized scores in 21st-century skills indicate that both gender and the educational intervention had a significant impact on student outcomes. The analysis, using both the Wilcoxon rank-sum test and the paired samples t-test, indicated statistically significant improvements in students' overall scores between the pretest and posttest.

For the Wilcoxon rank-sum test, the results revealed a significant difference between the overall scores at baseline and endline (W = 7829.500, p <.001, 95%CI -0.461 to -0.189). This suggests that there was a shift in the distribution of scores between the two time points, indicating that the students' overall 21st-century skills improved significantly after the intervention.

Similarly, the paired t-test results corroborate this finding. The test showed a statistically significant mean difference of -0.641 (t = -3.506, df = 237, p <.001, 95%CI -0.357 to -0.099). This indicates a significant increase in the overall score for 21st-century skills between the baseline and endline assessments, further confirming the positive impact of the intervention on students' skill development.

Table x. xxx.

Test type	Statistic	Z	Df	P	Effect Size	SE Effect Size	95% CI Lower	95% CI Upper
Student T-Test	-3.506		235	< .001	-0.228	0.078	-0.357	-0.099
Wilcoxon Test	7829.500	-4.229		< .001	-0.332	0.078	-0.461	-0.189

Note: For the Student T-test, effect size is given by Cohen's *d*. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.
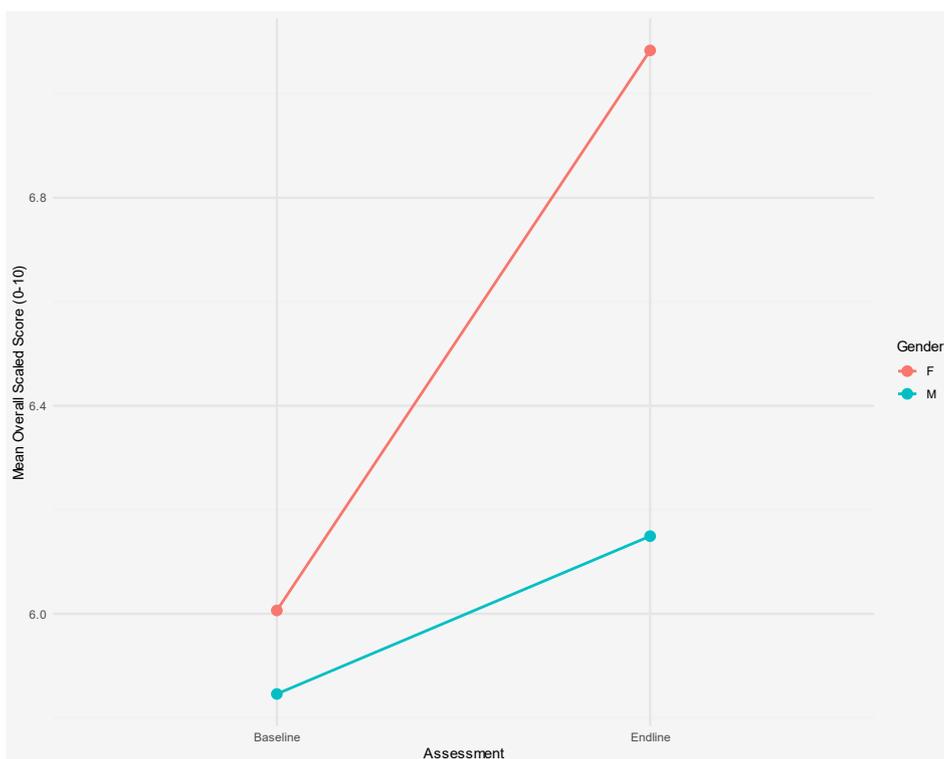

Figure 10. Comparison of Overall 21st Century Skills Scaled Scores by Gender.

The ANOVA results further confirm this, revealing significant main effects for both gender and measurement condition on the overall standardized scores. As shown in (Table 12), there was a significant main effect of gender on the overall scores, $F(1, 472) = 7.075$, $p = 0.008$. This result indicates that there were statistically significant differences in the overall performance between male and female students.

Table 12. Analysis of Variance (ANOVA) Results for the Effects of Gender and Measurement Condition on Overall Standardized Scores in 21st-Century Skills.

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gender	1	35.1	35.09	7.075	0.00808
Measurement Condition	1	49.0	48.96	9.874	0.00178
Gender:Measurement	1	17.5	17.51	3.532	0.06082
Residuals	472	2340.7	4.96		

Additionally, the analysis showed a significant main effect of measurement condition, $F(1, 472) = 9.874$, $p = 0.002$, suggesting that the intervention had a substantial impact on student performance, with overall scores differing significantly between the pretest and posttest conditions.

Furthermore, the interaction between gender and measurement condition approached significance, $F(1, 472) = 3.532$, $p = 0.061$. Although this interaction was not statistically significant at the conventional 0.05 level, it indicates a possible trend where the effect of the intervention on overall scores may vary between male and female students. This finding suggests that gender may play a role in how students respond to the intervention, warranting further investigation.

Figure 11 illustrates the changes in mean scores for the five 21st-century skills targeted by the intervention—Problem Solving, Collaborative Work, Creativity, Communication, and Critical Thinking—comparing baseline and endline assessments. Problem Solving saw an increase in the mean score from 6.40 at baseline to 6.72 at endline, accompanied by a slight rise in the standard error, indicating increased variability among students. Similarly, the mean score for Collaborative Work improved from 6.48 to 6.99, with a corresponding increase in standard error from 0.15 to 0.22, suggesting both better average performance and greater variation in outcomes.

In contrast, Creativity showed a decrease in the mean score from 6.78 at baseline to 6.13 at endline, while the standard error remained relatively stable, increasing slightly from 0.16 to 0.22. This suggests a decline in overall performance with a modest increase in variability. Communication demonstrated a significant improvement, with the mean score rising from 5.35 to 6.37, and an increase in standard error from 0.19 to 0.21, indicating better average performance and more diverse outcomes among students. Finally, Critical Thinking experienced a substantial increase in mean score, from 4.57 to 6.57, with the standard error also rising from 0.16 to 0.20, reflecting an overall enhancement in critical thinking skills alongside greater variability. Overall, these changes suggest that, except for Creativity, students improved in most assessed skills, though with varying degrees of increased variability.
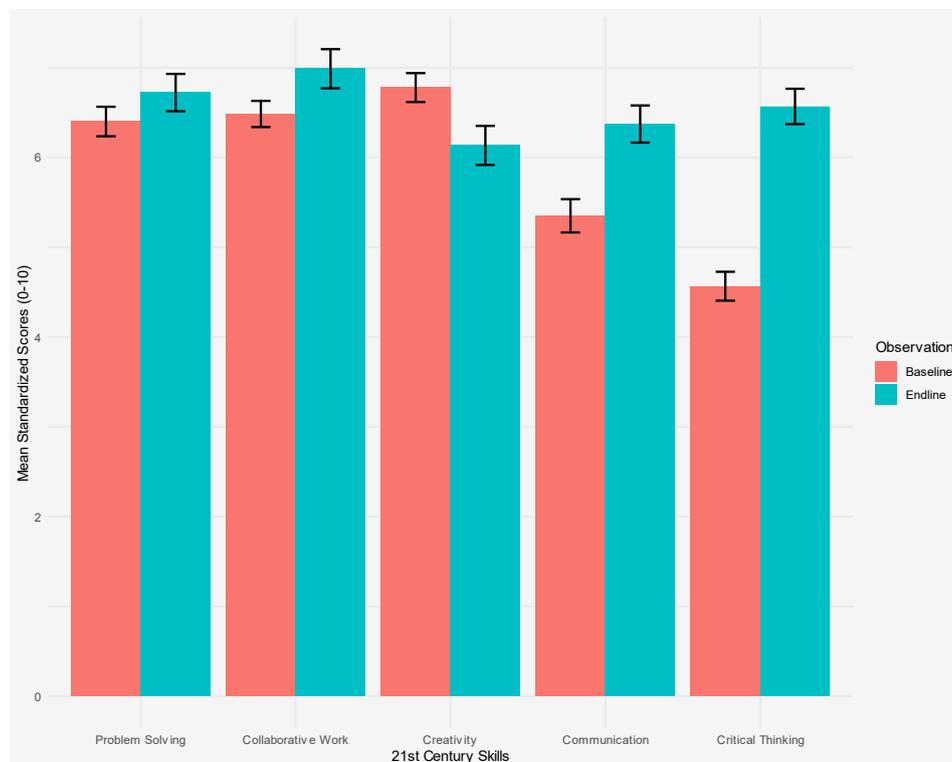
Figure 11. Assessment of Students' 21st-Century Skills, General Results by Competency (N=238).

5. Discussion

The findings of this study provide evidence that PBL significantly enhances students' competencies in language, mathematics, and science, while also fostering essential 21st-century skills.

The intervention's effectiveness is underscored by the substantial improvements observed in students' test scores from pretest to posttest assessments, as indicated by both paired t-tests and Wilcoxon rank-sum tests. These results align with previous literature that emphasizes the positive impact of PBL on academic achievement across multiple disciplines, including language and science education [47–49].

In the language component, the analysis revealed a notable increase in students' test scores, with a moderate effect size. The ANOVA results indicated that gender played a significant role in language performance, with girls generally outperforming boys. This finding is consistent with existing research that highlights gender differences in educational outcomes, suggesting that while PBL benefits all students, there may be variations in how different genders engage with and benefit from this instructional approach [50,51]. Importantly, the lack of a significant interaction between gender and measurement condition suggests that the improvements in language scores were uniformly distributed across genders, reinforcing the notion that PBL can be an equitable teaching strategy [52].

In mathematics, the intervention yielded significant improvements in test scores, corroborated by statistical analyses. The ANOVA results indicated that the type of test (pretest or posttest) significantly influenced math scores, affirming the effectiveness of PBL in enhancing mathematical abilities. The movement of many students from Beginning and In Progress levels to the Extended level signifies a robust positive shift in performance, echoing findings from other studies that demonstrate PBL's capacity to elevate students' mathematical competencies [47,49]. However, sustaining motivation and engagement throughout PBL activities remains a challenge, necessitating ongoing efforts to create stimulating learning environments [52].

The science component also exhibited significant improvements in test scores, with both paired t-tests and Wilcoxon rank-sum tests indicating meaningful changes. The ANOVA results further supported these findings, revealing a significant main effect of measurement condition on science scores. Interestingly, the impact of gender was not significant in this domain, suggesting that the intervention's effectiveness in enhancing scientific literacy was consistent across both male and female students. This aligns with the broader literature that suggests PBL can effectively engage students in science education, fostering inquiry and critical thinking skills [48,53].

The assessment of 21st-century skills provided additional insights into the intervention's impact. The ANOVA results revealed significant main effects of both gender and measurement condition on overall standardized scores, indicating that while the intervention was broadly effective, there may be underlying differences in how male and female students benefit from PBL. Notably, improvements were observed in Problem Solving, Collaborative Work, Communication, and Critical Thinking skills. However, the decrease in Creativity scores suggests that while PBL generally fosters critical and collaborative thinking, additional support may be necessary to effectively nurture creativity [28,54]. The increased variability in scores across most skills points to a diverse range of outcomes post-intervention, indicating that while average performance improved, the intervention may have had varying effects on different student groups [55].

In conclusion, the findings from this study underscore the potential of PBL to enhance student competencies and 21st-century skills, with significant gains observed across most areas. However, the variability in outcomes and specific challenges related to creativity highlight the need for further refinement of the intervention to ensure it addresses all aspects of student development effectively. Additionally, the observed gender differences suggest that tailored approaches may be necessary to maximize the benefits of PBL for all students, warranting further exploration into how these differences can be addressed in future implementations of PBL.

6. Limitations

This study employed a pretest-posttest single-group design, which, while valuable for assessing the impact of the educational intervention, comes with certain limitations. One of the primary limitations is the absence of a control group. The use of a control group would allow for a clearer attribution of observed changes in student performance to the intervention itself, rather than other external factors. However, in educational research, particularly when working with children, it is not

always feasible or ethical to withhold potentially beneficial interventions from a control group. The decision to implement the intervention across the entire group of participants was made to ensure that all students had access to the educational benefits it provided, in line with ethical considerations that prioritize the well-being and development of children.

Another limitation is the lack of random sampling, which affects the generalizability of the findings. The participants in this study were not randomly selected, meaning that the results may not be representative of the broader population. This non-randomized approach limits the extent to which the findings can be generalized beyond the specific group of students who participated in the intervention. However, despite this limitation, the study provides valuable insights into the potential impact of PBL on student competencies, specifically in math, language, and science, and 21st-century skills, offering a foundation for further research in more diverse settings.

In assessing 21st-century skills, the study relied on observation rubrics completed by a single observer. While rubrics are a useful tool for evaluating complex skills such as problem-solving, collaboration, and critical thinking, the use of only one observer introduces the risk of bias. The subjective nature of observations can lead to inconsistencies, and the absence of multiple observers prevents the calculation of inter-rater reliability (IRR), which would strengthen the validity of the assessment. Future research could benefit from incorporating more objective measures, such as performance-based assessments or technology-assisted evaluations, as well as involving multiple observers to enhance reliability. Running IRR analyses would further ensure the robustness of the findings, helping to mitigate the potential biases associated with single-observer assessments.

Despite these limitations, this study provides valuable insights into the effects of educational intervention on student learning outcomes. However, the limitations related to the study design, sampling, and assessment methods should be considered when interpreting the results. Future studies should address these limitations to provide a more comprehensive and generalizable understanding of the impact of educational interventions, particularly through the inclusion of control groups, random sampling, and more robust assessment methodologies.

7. Conclusions and Recommendations

The results of this study demonstrate the effectiveness of the Project-Based Learning (PBL) intervention in improving student competencies in language, mathematics, and science, as well as enhancing essential 21st-century skills among 3rd graders in rural Colombia. The intervention led to significant gains in standardized test scores and observable improvements in critical areas such as problem-solving, collaboration, communication, and critical thinking. These findings align with the growing body of literature that supports the use of PBL as a powerful instructional strategy that fosters both academic achievement and the development of key skills necessary for success in the 21st century.

The significant improvements observed in student performance, particularly the shift from lower to higher competency levels, indicate that the PBL approach effectively addressed the learning gaps exacerbated by the pandemic. The intervention not only helped students recover lost ground but also positioned them to excel in more advanced cognitive and practical tasks. The positive impact on 21st-century skills further underscores the value of integrating these competencies into the curriculum, ensuring that students are better equipped to navigate future challenges in education and beyond.

However, the study also revealed areas that require further attention. While most students showed improvement in various competencies, the decrease in creativity scores suggests that this area may require additional focus within PBL frameworks. Creativity is a critical component of 21st-century learning, and future iterations of the intervention should explore strategies to better support and enhance creative thinking among students. Additionally, the observed gender differences in overall standardized scores highlight the need for more tailored approaches that address the unique needs of male and female students within PBL environments.

Based on these findings, several recommendations can be made:

Firstly, enhance the focus on creativity. Given the observed decline in creativity scores, it would be beneficial for future interventions to include more targeted activities and assessments aimed at fostering creative thinking. Creativity is a critical component of 21st-century skills, essential for innovation and problem-solving. To address this, future iterations of the PBL intervention should incorporate more open-ended projects that encourage students to explore and express their ideas freely. Additionally, educators should be encouraged to use teaching strategies that promote divergent thinking, allowing students to approach problems from multiple perspectives and develop unique solutions.

Secondly, tailor interventions to address gender differences. The study's findings revealed significant gender differences in overall standardized scores, suggesting that male and female students may respond differently to PBL. To ensure that all students benefit equally from the intervention, future programs should consider gender-sensitive approaches. This could involve differentiated instruction that caters to the specific learning styles and needs of each gender, as well as providing gender-specific support strategies where necessary. Further research into the influence of gender on learning within PBL contexts is also recommended to better understand these dynamics and to inform the development of more inclusive educational practices.

Thirdly, expand the use of PBL in rural education. The success of this intervention in rural Colombia indicates that PBL is a viable and effective approach for improving educational outcomes in underserved areas. Given the challenges faced by students in rural regions, including limited access to resources and educational support, the expansion of PBL could provide significant benefits. Policymakers and educators should consider implementing PBL on a broader scale across rural schools, ensuring that students in these areas have access to high-quality, engaging, and relevant educational experiences. This expansion should be supported by ongoing training for teachers and administrators to effectively implement and sustain PBL practices.

Fourthly, incorporate more objective assessment methods. To strengthen the evaluation of 21st-century skills, it is recommended that future studies include more objective measures alongside observation rubrics. While rubrics provide valuable insights, their subjective nature can introduce bias, particularly when completed by a single observer. Incorporating performance-based assessments, technology-assisted tools, and involving multiple observers can enhance the reliability and validity of the findings. Additionally, conducting inter-rater reliability (IRR) analyses would ensure consistency in the assessment process and contribute to the robustness of the results.

Finally, conduct longitudinal studies to better understand the long-term impact of PBL on student learning and development. While this study provides valuable insights into the immediate effects of the intervention, it is essential to track students over time to assess how PBL influences their academic trajectories, skill development, and future success. Longitudinal studies would provide a more comprehensive understanding of the sustained impact of PBL, allowing educators and policymakers to refine and adapt the approach to maximize its benefits for students in diverse educational contexts.

consequences. All data collected during the intervention were anonymized to protect the privacy and confidentiality of the participants.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. UNICEF *Education in Latin America and the Caribbean at a Crossroads. Regional Monitoring Report SDG4 - Education 2030*; UNESCO: Paris, France, 2022;

2. Icfes *Programa Para La Evaluación Internacional de Alumnos (PISA). Informe Nacional de Resultados Para Colombia 2022*; Icfes: Bogotá, 2024;

3. P21 Framework for 21st Century Learning 2009.

4. Wu, Y.; Wahab, A. Research on the Material and Human Resource Allocation in Rural Schools Based on Social Capital Theory. *International Journal of Education and Humanities* **2023**, *11*, 46–49, doi:10.54097/ijeh.v11i3.14451.

5. Freeman, S.; Eddy, S.L.; McDonough, M.; Smith, M.K.; Okoroafor, N.; Jordt, H.; Wenderoth, M.P. Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences* **2014**, *111*, 8410–8415, doi:10.1073/pnas.1319030111.

6. Prince, M. Does Active Learning Work? A Review of the Research. *J of Engineering Edu* **2004**, *93*, 223–231, doi:10.1002/j.2168-9830.2004.tb00809.x.

7. Vergara, D.; Paredes-Velasco, M.; Chivite, C.; Fernández-Arias, P. The Challenge of Increasing the Effectiveness of Learning by Using Active Methodologies. *Sustainability* **2020**, *12*, 8702, doi:10.3390/su12208702.

8. Beckett, G. Teacher and Student Evaluations of Project-Based Instruction. *TESL* **2002**, *19*, 52, doi:10.18806/tesl.v19i2.929.

9. Guslyakova, A.; Guslyakova, N.; Valeeva, N.; Veretennikova, L. Project-Based Learning Usage in L2 Teaching in a Contemporary Comprehensive School (on the Example of English as a Foreign Language Classroom). *revtee* **2021**, *14*, e16754, doi:10.20952/revtee.v14i33.16754.

10. Haatainen, O.; Aksela, M. Project-Based Learning in Integrated Science Education: Active Teachers' Perceptions and Practices. *LUMAT* **2021**, *9*, doi:10.31129/LUMAT.9.1.1392.

11. Padmadewi, N.N.; Suarcaya, P.; Artini, L.P.; Munir, A.; Friska, J.; Husein, R.; Paragae, I. Incorporating Linguistic Landscape into Teaching: A Project-Based Learning for Language Practices at Primary School. *International j.of Elementary Education* **2023**, *7*, 467–477, doi:10.23887/ijee.v7i3.58792.

12. Yamil, K.; Said, N.; Swanto, S.; Din, W.A. Enhancing ESL Rural Students' Speaking Skills Through Multimedia Technology-Assisted Project-Based Learning: Conceptual Paper. *IJEPC* **2022**, *7*, 249–262, doi:10.35631/IJEPC.748018.

13. Sagita, L.; Putri, R.I.I.; Zulkardi; Prahmana, R.C.I. Promising Research Studies between Mathematics Literacy and Financial Literacy through Project-Based Learning. *j. math. educ.* **2023**, *13*, 753–772, doi:10.22342/jme.v13i4.pp753-772.

14. Ndiung, S.; Menggo, S. Project-Based Learning in Fostering Creative Thinking and Mathematical Problem-Solving Skills: Evidence from Primary Education in Indonesia. *IJLTER* **2024**, *23*, 289–308, doi:10.26803/ijlter.23.8.15.

15. Lavonen, L.; Loukomies, A.; Vartianen, J.; Palojoki, P. Promoting 3rd Grade Pupils' Learning of Science Knowledge through Project-Based Learning in a Finnish Primary School. *NorDiNa* **2023**, *19*, 181–199, doi:10.5617/nordina.9686.

16. Jabr, E.A.H.A.-S.; Zaki, S.B.Prof.Dr.S.Y.; Mohamed, Prof.Dr.M.A.S.; Ahmed, Assist.Prof.D.M.F. The Effectiveness of Curriculum Developer in Science in View of Project Based Learning to Developing Problem Solving Skills for Primary Stage Pupils. *Educational Research and Innovation Journal* **2023**, *3*, 97–131, doi:10.21608/erji.2023.311742.

17. Rijken, P.E.; Fraser, B.J. Effectiveness of Project-Based Mathematics in First-Year High School in Terms of Learning Environment and Student Outcomes. *Learning Environ Res* **2024**, *27*, 241–263, doi:10.1007/s10984-023-09477-7.

18. TeGrootenhuis, J. The Effectiveness of Project-Based Learning in the Science Classroom. Master's thesis, Northwestern College, Iowa: Orange City, 2018.

19. Zhang, L.; Ma, Y. A Study of the Impact of Project-Based Learning on Student Learning Effects: A Meta-Analysis Study. *Front. Psychol.* **2023**, *14*, 1202728, doi:10.3389/fpsyg.2023.1202728.

20. Fadhil, M.; Kasli, E.; Halim, A.; Evendi; Mursal; Yusrizal Impact of Project Based Learning on Creative Thinking Skills and Student Learning Outcomes. *J. Phys.: Conf. Ser.* **2021**, *1940*, 012114, doi:10.1088/1742-6596/1940/1/012114.

21. Biazus, M. de O.; Mahtari, S. The Impact of Project-Based Learning (PjBL) Model on Secondary Students' Creative Thinking Skills. *International Journal of Essential Competencies in Education* **2022**, *1*, 38–48, doi:10.36312/ijece.v1i1.752.

22. Valtonen, T.; Sointu, E.; Kukkonen, J.; Kontkanen, S.; Lambert, M.C.; Mäkitalo-Siegl, K. TPACK Updated to Measure Pre-Service Teachers' Twenty-First Century Skills. *Australasian Journal of Educational Technology* **2017**, *33*, doi:10.14742/ajet.3518.

23. Anagün, Ş.S. Teachers' Perceptions about the Relationship between 21st Century Skills and Managing Constructivist Learning Environments. *INT J INSTRUCTION* **2018**, *11*, 825–840, doi:10.12973/iji.2018.11452a.

24. Herfina, E.S. Feasibility Test of 21st Century Classroom Management Through Development Innovation Configuration Map. *Jurnal Pendidikan dan Pengajaran Guru Sekolah Dasar (JPPGuseda)* **2022**, *5*, 101–104, doi:10.55215/jppguseda.v5i3.6510.

25. Fitria, D.; Lufri, L.; Elizar, E.; Amran, A. 21st Century Skill-Based Learning (Teacher Problems In Applying 21st Century Skills). *IJHESS* **2023**, *2*, doi:10.55227/ijhess.v2i4.409.

26. Varghese, J.; Musthafa, undefined M.N.M.A. Investigating 21st Century Skills Level among Youth. *GiLE Journal of Skills Development* **2021**, *1*, undefined-undefined, doi:10.52398/gjsd.2021.v1.i2.pp99-107.

27. Bell, S. Project-Based Learning for the 21st Century: Skills for the Future. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas* **2010**, *83*, 39–43, doi:10.1080/00098650903505415.

28. Usmeldi, U. The Effect of Project-Based Learning and Creativity on the Students' Competence at Vocational High Schools. In Proceedings of the Proceedings of the 5th UPI International Conference on Technical and Vocational Education and Training (ICTVET 2018); Atlantis Press: Bandung, Indonesia, 2019.

29. Saputra, I.G.N.H.; Joyoatmojo, S.; Harini, H. The Implementation of Project-Based Learning Model and Audio Media Visual Can Increase Students' Activities. *International Journal of Multicultural and Multireligious Understanding* **2018**, *5*, 166–174, doi:10.18415/ijmmu.v5i4.224.

30. Ana, A.; Subekti, S.; Hamidah, S. The Patisserie Project Based Learning Model to Enhance Vocational Students' Generic Green Skills.; Atlantis Press, February 2015; pp. 24–27.

31. Mislevy, R.J.; Riconscente, M. Evidence-Centered Assessment Designs: Layers, Concepts and Terminology. In *Handbook of test development*; Downing, S., Haladyna, T., Eds.; Earlbaum: Mahwah, NJ, 2006; pp. 61–90.

32. *Bayesian Networks in Educational Assessment*; Almond, R.G., Mislevy, R.J., Steinberg, L.S., Yan, D., Williamson, D., Eds.; Statistics for social and behavorial sciences; Springer: New York, NY, 2015; ISBN 978-1-4939-2124-9.

33. Mislevy, R.J.; Haertel, G.D. Implications of Evidence-Centered Design for Educational Testing. *Educational Measurement: Issues and Practice* **2007**, *25*, 6–20, doi:10.1111/j.1745-3992.2006.00075.x.

34. Mislevy, R.J.; Almond, R.G.; Lukas, J.F. *A Brief Introduction to Evidence-Centered Design*; University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST): Los Angeles, 2004;

35. Ministerio de Educación *Estándares Básicos de Competencias*; MEN: Bogotá, Colombia, 2006;

36. Webb, N.L. Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. Research Monograph No. 6. 1997.

37. Webb, N.L. Identifying Content for Student Achievement Tests. In *Handbook of test development*; Downing, S., Haladyna, T., Eds.; LEA: London, 2011; pp. 155–180.

38. Webb, N.L. Web Alignment Tool 2005.

39. Jonsson, A.; Svingby, G. The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review* **2007**, *2*, 130–144, doi:10.1016/j.edurev.2007.05.002.

40. Weiss, C. Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*; Connell, J.P., Kubisch, A.C., Schorr, L., Weiss, C., Eds.; Aspen Institute: Washington, DC, 1995; pp. 65–92.

41. Weiss, C. *Evaluation : Methods for Studying Programs and Policies*; 2nd ed.; Prentice Hall: Upper Saddle River N.J., 1998; ISBN 978-0-13-309725-2.

42. Staff, T. A Depth Of Knowledge Rubric For Reading, Writing, And Math. *TeachThought* 2014.

43. Limesurvey GmbH LimeSurvey: An Open Source Survey Tool 2024.

44. McDowell, I. *Measuring Health: A Guide to Rating Scales and Questionnaires*; 2006;

45. De Vaus, D.A. *Surveys in Social Research*; Routledge: New York, NY, 2014; ISBN 978-0-415-53015-6.

46. R Core Team *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022;

47. Craig, T.T.; Marshall, J. Effect of Project-based Learning on High School Students' State-mandated, Standardized Math and Science Exam Performance. *J Res Sci Teach* **2019**, *56*, 1461–1488, doi:10.1002/tea.21582.
48. Kokotsaki, D.; Menzies, V.; Wiggins, A. Project-Based Learning: A Review of the Literature. *Improving Schools* **2016**, *19*, 267–277, doi:10.1177/1365480216659733.
49. Uluçınar, U. The Effect of Problem-Based Learning in Science Education on Academic Achievement: A Meta-Analytical Study. *SEI* **2023**, *34*, 72–85, doi:10.33828/sei.v34.i2.1.
50. Alenezi, A. Using Project-Based Learning Through the Madrasati Platform for Mathematics Teaching in Secondary Schools: *International Journal of Information and Communication Technology Education* **2023**, *19*, 1–15, doi:10.4018/IJICTE.332372.
51. Kurniadi, D.; Cahyaningrum, I.O. Exploring Project-Based Learning for Young Learners in English Education. *NextGen* **2023**, *1*, 10–21, doi:10.58660/nextgen.v1i2.35.
52. Fang, C.Y.; Zakaria, M.I.; Iwani Muslim, N.E. A Systematic Review: Challenges in Implementing Problem-Based Learning in Mathematics Education. *IJARPED* **2023**, *12*, Pages 1261-1271, doi:10.6007/IJARPED/v12-i3/19158.
53. Kurt, G.; Akoglu, K. Project-Based Learning in Science Education: A Comprehensive Literature Review. *INTERDISCIP J ENV SCI ED* **2023**, *19*, e2311, doi:10.29333/ijese/13677.
54. Usmeldi, U.; Amini, R. Creative Project-Based Learning Model to Increase Creativity of Vocational High School Students. *IJERE* **2022**, *11*, 2155, doi:10.11591/ijere.v11i4.21214.
55. Belwal, R.; Belwal, S.; Sufian, A.B.; Al Badi, A. Project-Based Learning (PBL): Outcomes of Students' Engagement in an External Consultancy Project in Oman. *ET* **2020**, *63*, 336–359, doi:10.1108/ET-01-2020-0006.