Revista Multidisciplinaria Perspectivas Investigativas
Multidisciplinary Journal Investigative Perspectives
Vol. 6(4), 27-34, 2026
https://doi.org/10.62574/rmpi.v6i4.525

Problem-based learning with ChatGPT and understanding of kinematics in tenth grade

Aprendizaje basado en problemas con ChatGPT y comprensión de la cinemática en décimo grado

Jhonn Edisson Cruz-Guerra
Jhonn-e.cruz-g@up.ac.pa
Universidad de Panamá, Panamá, Provincia de Panamá, Panamá
https://orcid.org/0009-0004-1838-1785

ABSTRACT

The study aims to assess the incidence of ChatGPT, integrated into problem-based learning (PBL), on the understanding of kinematics among tenth-grade students of a technical-industrial school in Chiquinquirá, Colombia. A mixed approach with a quasi-experimental design was used, comparing an experimental group and a control group through pretest, posttest, observation diaries, validation matrices and interviews. The results show that the conceptual score (maximum 18) rose from 5.35 to 11.71 in the experimental group and from 3.27 to 6.39 in the control group; interpretation of the position-time slope advanced from 2.9% to 50.0% in the former, and difficulties persisted with the area under the velocity-time curve. It is concluded that the strategy was associated with a greater descriptive change in conceptual performance, without allowing causal attribution, and that it required teacher mediation and critical validation of the responses.

Descriptors: artificial intelligence; teaching methods; physics. (Source: UNESCO Thesaurus).

RESUMEN

El estudio se propone valorar la incidencia de ChatGPT, integrado al aprendizaje basado en problemas (ABP), en la comprensión de la cinemática en estudiantes de grado décimo de una institución técnico industrial de Chiquinquirá, Colombia. Se empleó un enfoque mixto con diseño cuasiexperimental, que comparó un grupo experimental y un grupo control mediante pretest, postest, diarios de observación, matrices de validación y entrevistas. Los resultados muestran que el puntaje conceptual (máximo 18) pasó de 5,35 a 11,71 en el grupo experimental y de 3,27 a 6,39 en el control; la interpretación de la pendiente posición-tiempo avanzó de 2,9 % a 50,0 % en el primero, y persistieron dificultades con el área bajo la curva velocidad-tiempo. Se concluye que la estrategia se asoció con un mayor cambio descriptivo del desempeño conceptual, sin permitir una atribución causal, y que requirió mediación docente y validación crítica de las respuestas.

Descriptores: inteligencia artificial; métodos de enseñanza; física. (Fuente: Tesauro UNESCO).

Received: 5/10/2026. Reviewed: 12/10/2026. Accepted: 18/10/2026. Published: 5/11/2026.
Research Articles Section

INTRODUCTION

Understanding kinematics is one of the most persistent obstacles in physics teaching at secondary level, since it requires coordinating verbal, graphical and algebraic representations of quantities such as position, velocity and acceleration, while overcoming intuitive conceptions rooted in everyday experience. Klein, et al., (2021) validated the item objectives of the test of understanding graphs in kinematics by tracking eye movements, which places graph reading as a specific and assessable skill within the learning of motion. Lecture-based methods that favour the repetition of formulas therefore have limited power to change the way students explain phenomena, and strategies that position them as active builders of their own understanding need to be explored.

Among active methodologies, problem-based learning (PBL) has empirical support in physics teaching. Becerra-Labra, et al., (2012) documented that a problem-based structure of contents improves both conceptual learning and problem-solving ability, while Argaw, et al., (2017) reported favourable effects of PBL instruction on motivation and on physics problem-solving skills. The systematic review by Nicholus, et al., (2023) describes the role of this approach in teaching and learning the discipline, although the evidence is not uniform: Marcinauskas, et al., (2024), comparing PBL with traditional teaching among engineering students, observed that the former fostered teamwork, the presentation of results and critical thinking, whereas traditional teaching was associated with better marks in closed-question exams. At the regional level, Díaz-Santamaría, et al., (2017) analysed project-based learning in the teaching of kinematics in Ecuador, which shows an interest in approaching this content through active methods.

At the same time, the arrival of language models such as ChatGPT opened up new possibilities and new risks for physics education. Liang, et al., (2023) explored the potential of this tool in teaching the discipline, and Bitzenbauer (2023), in a pilot study with 53 secondary school students, tested two easy-to-implement activities aimed at critical thinking, after which students' perception of the tool improved. Polverini & Gregorcic (2024), for their part, assessed ChatGPT-4 with the test of understanding graphs in kinematics and found that its average performance was similar to that of high school students, with strengths in reasoning and in proposing solution strategies but with difficulties in reading graphs accurately, which is why they recommend using it with caution and critical judgement.

However, evidence that combines PBL with ChatGPT through explicit validation protocols, in technical secondary schools serving low socioeconomic strata, is scarce in the sources reviewed, and the available studies tend to describe potential rather than compare groups. Muñoz-Alvarez, et al., (2025) documented PBL as a strategy for teaching physics in Chilean technical-professional higher education, which supports its relevance in technical training settings, although its combination with generative artificial intelligence tools requires evidence of its own.

Building on the above, and as a continuation of the author's previous documentary review on the convergence between artificial intelligence and PBL in understanding kinematics, this article, derived from a completed doctoral thesis, aims to assess the incidence of the artificial intelligence ChatGPT, applied through problem-based learning, on the understanding of kinematics among tenth-grade students at the Institución Educativa Técnico Industrial “Julio Flórez” in Chiquinquirá, Boyacá, Colombia.

METHOD

The research was framed within the phenomenological-hermeneutic paradigm and adopted a mixed approach with a descriptive quasi-experimental design, characterised by partial control of variables and the absence of random assignment, since the groups were formed from pre-existing cohorts (Creswell & Creswell, 2018). An experimental group, which worked with the integration of artificial intelligence and PBL, was therefore compared with a control group that received conventional teaching; quantitative and qualitative data were collected and analysed concurrently so that the results could be cross-checked through triangulation.

The study was carried out at the Institución Educativa Técnico Industrial “Julio Flórez” in Chiquinquirá, Boyacá, whose students come mostly from socioeconomic strata 0, 1 and 2. Participants were selected purposively from the four tenth-grade groups: class 10-1, with 34

students, served as the control group, and class 10-2, with 35 students, as the experimental group. Regarding ethical aspects, institutional approval and informed consent from parents or guardians were obtained, participation was voluntary, and records of interaction with the artificial intelligence were kept anonymised. Depending on the questionnaires available, the analysis included 33 pretests from the control group and 34 from the experimental group, as well as 33 posttests from the control group and 28 from the experimental group.

Both groups covered the same kinematics content over approximately eight weeks, with three hours per week. The control group worked through lectures, printed guides, textbook exercises and teacher-led discussion, without using ChatGPT or formally applying PBL. The experimental group followed the eight phases proposed by Morales & Landa (2004), namely problem identification, analysis, hypothesis, information search, synthesis, application, evaluation and reflection, in line with the conception of Barrows (1986), which takes the problem as the starting point for acquiring and integrating new knowledge, and used ChatGPT as scaffolding through role prompts that asked for hints rather than the final answer. Every generated response had to pass a six-step validation protocol: recording the prompt and the response, identifying variables, manual calculation, dimensional analysis, conceptual comparison with class notes and the teacher's explanation, and a critical decision to accept, correct or refute it. Before projectile motion, a brief scaffold on trigonometry and vector decomposition was provided, and four sessions of the experimental group were recorded: free fall (04/08/2026), half-projectile launch (11/08/2026), projectile motion (01/09/2026) and integrated problem solving (08/09/2026).

The instruments were a 13-question pretest, organised into prior knowledge (Q1 to Q5), understanding of concepts (Q6 to Q9) and learning processes (Q10 to Q13), and a 20-question posttest that kept those categories and added the process of apprehension (Q14 to Q20), understood as the meaningful appropriation of knowledge, that is, the ability to make a concept one's own, explain it in one's own words, relate it to everyday life and use it in new problems. Questions 1 to 9 and 14 to 20 were scored with a three-level rubric (0, 1 and 2 points), with a maximum of 18 points in the common block and 14 in the apprehension block; questions 10 to 13 were analysed qualitatively. These instruments were complemented by ten transcripts of semi-structured interviews, the pedagogical observation diaries of the four sessions and the group matrices for analysing and validating the artificial intelligence's responses, nine of the ten groups providing usable information.

Scores were described using mean, standard deviation and median, and were compared between groups with the Mann-Whitney U test, together with the rank-biserial correlation as a measure of effect size, given the ordinal nature of the rubric. No paired test was applied and no individual gain was calculated, since a verified match between the questionnaires of both moments was not available; the change between pretest and posttest was therefore expressed as the difference between group means. The qualitative analysis organised interviews, diaries and matrices into planned and emerging categories, and triangulation contrasted the sources without making up for the limitations of the design.

RESULTS

The results are presented below in four sections: conceptual performance before and after the intervention, the apprehension process, the implementation and use of ChatGPT, and students' perceptions together with the triangulation of sources.

Conceptual performance in the pretest and posttest

In the pretest, both groups started with limited conceptual mastery: the control group obtained a mean of 3.27 points out of 18 and the experimental group 5.35, a statistically detectable difference before the intervention ($U = 326.5$; $p = .003$; $r = .42$). After the eight weeks, the mean of the control group was 6.39 and that of the experimental group 11.71 ($U = 105$; $p < .001$; $r = .77$), as summarised in Table 1.

Table 1. *Conceptual score (questions 1 to 9, maximum 18) in the pretest and posttest by group*.

Group	n pretest	M (SD) pretest	n posttest	M (SD) posttest	Mean change
Control	33	3.27 (2.11)	33	6.39 (3.29)	+3.12
Experimental	34	5.35 (2.92)	28	11.71 (3.11)	+6.36

Note: developed in-house. M = mean; SD = standard deviation. The mean change is the

difference between group means, not a paired individual gain.

In relation to the maximum score, the average of the control group rose from 18.2% to 35.5% and that of the experimental group from 29.7% to 65.1%, so the difference between the two mean changes was 3.24 points in favour of the condition with artificial intelligence and PBL. By category, in prior knowledge (maximum 10) the control group advanced from 2.82 to 4.21 points and the experimental group from 3.91 to 6.96; in understanding of concepts (maximum 8), the control group went from 0.45 to 2.18 and the experimental group from 1.44 to 4.75.

The change was not even across contents (Table 2). In the experimental group, adequate responses, corresponding to level 2 of the rubric, rose from 8.8% to 85.7% in the definition of kinematics, from 2.9% to 50.0% in the interpretation of the position-time slope, from 26.5% to 67.9% in uniform rectilinear motion, from 5.9% to 46.4% in speed and velocity, and from 5.9% to 39.3% in the possibility of motion with zero acceleration. The control group also advanced, especially in types of motion (from 39.4% to 75.8%) and in accelerated motion (from 3.0% to 24.2%), but reached only 6.1% in the position-time slope. Types of motion showed equivalent percentages in the posttest (75.8% and 75.0%).

Table 2. *Percentage of conceptually adequate responses in the pretest and posttest by concept*.

Concept	Control pretest	Control posttest	Experimental pretest	Experimental posttest
Kinematics	0.0%	36.4%	8.8%	85.7%
Motion	0.0%	0.0%	2.9%	7.1%
Speed and velocity	0.0%	6.1%	5.9%	46.4%
Types of motion	39.4%	75.8%	52.9%	75.0%
Uniform rectilinear motion	9.1%	18.2%	26.5%	67.9%
Position-time slope	0.0%	6.1%	2.9%	50.0%
Accelerated motion	3.0%	24.2%	41.2%	64.3%
Zero acceleration with motion	0.0%	9.1%	5.9%	39.3%
Mass and free fall	3.0%	21.2%	2.9%	25.0%

Note: developed in-house. A response scored 2 points on the rubric was considered adequate. Variations are group-level and descriptive.

By contrast, two contents kept low percentages in both groups: the general concept of motion, which required integrating change of position, frame of reference and time (0.0% and 7.1% in the posttest), and the relationship between mass and free fall (21.2% in the control group and 25.0% in the experimental group), with responses that associated greater mass with a faster fall.

Apprehension process

In questions 14 to 20, applied only in the posttest, the control group obtained a mean of 3.33 points out of 14 (SD = 2.42) and the experimental group 7.11 (SD = 3.17), a statistically detectable difference ($U = 164.0$; $p < .001$; $r = .65$), in a block with no equivalent measurement in the pretest. As Table 3 shows, the experimental group outperformed the control group in six of the seven contents, with the largest difference in the graphical representation of motion (60.7% versus 0.0%); in contrast, the interpretation of the area under the velocity-time curve was achieved by only one student in each group (3.0% and 3.6%).

Table 3. *Percentage of adequate responses in the posttest apprehension items (Q14 to Q20)*.

Content assessed	Control	Experimental
Average and instantaneous velocity (Q14)	9.1%	35.7%
Projectile motion (Q15)	30.3%	57.1%
Transition from accelerated motion to URM (Q16)	6.1%	28.6%
Sign of the acceleration due to gravity (Q17)	36.4%	64.3%
Area under the velocity-time curve (Q18)	3.0%	3.6%
Distance and displacement (Q19)	6.1%	32.1%
Graphical representation of motion (Q20)	0.0%	60.7%

Note: developed in-house. Percentages correspond to responses scored 2 points. URM = uniform rectilinear motion.

Implementation and use of ChatGPT

The diaries of the four sessions show a shift from a use of ChatGPT focused on immediate checking towards more specific queries, linked to collaborative work and the teacher's guidance.

In the first session, difficulties in interpreting problem statements, identifying variables and handling units prevailed; in the second, groups shared out roles and began to detect inconsistencies in the tool's answers; in the third, they combined peer discussion, teacher guidance and consultation with the artificial intelligence; and in the fourth, several students used it to understand or review procedures and not only to obtain the final result. Attendance ranged between 30 and 32 students.

Of the ten planned groups, nine completed the validation matrix, in which three common uses were identified: checking procedures previously solved by hand, asking for guidance on specific doubts and critically reviewing the generated answers. Six of the nine records noted at least one limitation, inconsistency or omission attributed to the tool, such as skipped steps, incomplete procedures or formulas without explanation; groups 1 and 3 rejected answers they considered incomplete or potentially confusing, and group 8 recorded corrections of units and of the sign of gravity suggested by the tool.

Difficulties with rearranging equations, unit conversion, signs and frames of reference persisted into the advanced sessions, as did the initial tendency to expect direct solutions. It was also documented that ChatGPT marked as incorrect a procedure that it later acknowledged as correct when questioned by the students. Records from some groups were incomplete, and group 10 did not provide enough information for the validation analysis.

Student perceptions and triangulation

The ten interview transcripts describe artificial intelligence as explanatory and procedural support: “It helped me understand the variables, how to use them, motion and, above all, understand the equations” (Interviewee 1). PBL was associated with order and a sequence of steps (“I relied on PBL, putting things in order so I could solve the exercise”, Interviewee 10), and verification with one's own calculation and the teacher (“First we did the exercise and then we asked the teacher whether it was right”, Interviewee 1). Participants valued the availability of the tool outside class and the chance to ask questions they felt embarrassed to put to the teacher, and they proposed criteria for responsible use: “You don't just follow what the AI says, you have to interpret what ChatGPT is doing” (Interviewee 6). Since the transcripts identify the interviewees as students of class 10-1, while the design distinguishes the two conditions, no qualitative comparisons between groups were made.

Table 4 brings together, by axis of analysis, the quantitative, process and perception evidence obtained in the study.

Table 4. *Quantitative, process and perception evidence by axis of analysis*.

Axis of analysis	Quantitative evidence	Process and perception evidence
Conceptual understanding	Block Q1-Q9: experimental group from 5.35 to 11.71; control group from 3.27 to 6.39.	More specific queries and review of procedures in the diaries; explanatory support described in the interviews.
Graphical interpretation	Position-time slope from 2.9% to 50.0% and 60.7% in graphical representation (Q20) in the experimental group.	Explanations and images valued by students; area under the velocity-time curve at 3.6%.
Critical verification	No direct quantitative measure.	Matrices with review of variables, units and the decision to accept, correct or reject; comparison with the teacher and peers in the interviews.
Autonomy and mediation	No direct quantitative measure.	Greater group organisation and own decisions about what to ask and what to accept; the teacher as the reference for validation.

Note: developed in-house from the pretest, posttest, observation diaries, validation matrices and interviews.

DISCUSSION

The results indicate that integrating ChatGPT into a PBL sequence was associated with a greater descriptive change in the conceptual performance of tenth-grade students, since the experimental group raised its mean by 6.36 points against 3.12 in the control group and reached higher proportions of adequate responses in most concepts and in six of the seven apprehension contents. Consistent with this interpretation, Becerra-Labra, et al., (2012) and Argaw, et al., (2017) documented effects of PBL on conceptual learning and physics problem solving, the synthesis by Nicholus, et al., (2023) supports its role in the discipline, and Kanyesigye, et al., (2022) showed

that problem-based instruction made it possible to remedy entrenched conceptual difficulties in the study of mechanical waves. However, the design does not allow the contribution of artificial intelligence to be isolated from that of PBL, collaborative work and teacher mediation, so the difference should be read as an association and not as a causal effect. Triangulation showed convergence between the greater descriptive change in the experimental group, the progression observed in the classroom and the perceived usefulness of the tool, although the autonomy recorded was guided by the pedagogical design and the persistent difficulties rule out speaking of an even mastery of kinematics.

One of the most visible changes occurred in graphical interpretation, as identifying the position-time slope as velocity rose from 2.9% to 50.0% in the experimental group, and 60.7% correctly represented the transition from constant acceleration to constant velocity. Since Klein, et al., (2021) place graph reading as a specific skill in kinematics, this progress is relevant; Polverini & Gregorcic (2024), in turn, warn that ChatGPT-4 reasons well and proposes solution strategies but reads graphs with difficulty, so the improvement observed can hardly be attributed to the tool's graph reading and is probably linked to the explanations, work with peers and teacher guidance, an aspect that needs to be tested in later studies. The persistence of difficulties with the area under the velocity-time curve also indicates that integrating graphical representations with abstract quantities calls for a specific intervention.

As for the validation of responses, the protocol worked as a central component of the strategy: the matrices and interviews show that ChatGPT's answers were not accepted automatically but were checked against manual calculations, units, signs and the teacher's judgement. This practice is in line with the caution recommended by Polverini & Gregorcic (2024) and with the orientation towards critical thinking about the tool that Bitzenbauer (2023) proposed in his activities. Nevertheless, the matrices were designed precisely to prompt review, so they do not show that students carried it out spontaneously outside the activity, and the records reveal that the level of detail in verification varied across groups.

The results should be qualified with the evidence of Marcinauskas, et al., (2024), who found that PBL fostered teamwork and critical thinking, whereas traditional teaching was associated with better marks in closed-question exams. In the present study, the rubric favoured open answers that required explaining physical relationships, so the results could vary with instruments of another kind; along these lines, types of motion, an item that mainly assesses recognition of names, showed no differences between groups in the posttest.

Teacher mediation and collaborative work remained constant reference points, both in the diaries and in the interviews, which suggests that the tool worked within a network of human support and not as a substitute for the teacher. Similarly, Díaz-Santamaría, et al., (2017) analysed the teaching of this content through projects in Ecuador and Muñoz-Alvarez, et al., (2025) documented the relevance of PBL in technical-professional training, so the experience described here adds to this line of work, with the particular feature of incorporating a generative artificial intelligence tool and a validation protocol. In a school with students from strata 0, 1 and 2, the availability of the tool outside class hours widened the opportunities to ask questions, although availability is not the same as cognitive autonomy.

The limitations include the non-random formation of the groups, the initial advantage of the experimental group, the lack of individual matching between pretest and posttest, the loss of six posttest questionnaires from the experimental group (17.6%), the group-level attribution of the interviews, which prevents a qualitative comparison of the conditions, and the uneven nature of the validation records. Looking ahead, studies with random assignment or individual matching, retention measurements and designs that separate the effect of PBL from that of artificial intelligence are needed.

CONCLUSIONS

The integration of ChatGPT into a PBL sequence, accompanied by teacher mediation and an explicit validation protocol, was associated with a greater descriptive increase in the conceptual performance in kinematics of tenth-grade students: the experimental group went from 5.35 to 11.71 points out of 18 and the control group from 3.27 to 6.39, with visible progress in the interpretation of the position-time slope, the graphical representation of motion, uniform rectilinear

motion and the distinction between speed and velocity. The results were not uniform and do not allow causal attribution, since the groups were not equivalent at the outset and difficulties persisted with the concept of motion, the relationship between mass and free fall, and the area under the velocity-time curve. The contribution of the study therefore lies in showing that the value of artificial intelligence in teaching kinematics depends on its pedagogical integration: when subjected to verification, discussion and justification, it widens the opportunities to explain, organise and correct problem solving, without replacing the student's reasoning.

FINANCING

Non-monetary.

CONFLICT OF INTEREST

There is no conflict of interest with any individuals or institutions related to this research.

ACKNOWLEDGMENTS

To the tenth-grade students of the Institución Educativa Técnico Industrial “Julio Flórez” in Chiquinquirá, for their participation, and to Dr. César Ignacio Báez Quintero, supervisor of the doctoral thesis from which this article derives.

REFERENCES

Argaw, A. S., Haile, B. B., Ayalew, B. T., & Kuma, S. G. (2017). The effect of problem based learning (PBL) instruction on students' motivation and problem solving skills of physics. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(3), 857–871. https://doi.org/10.12973/eurasia.2017.00647a

Barrows, H. S. (1986). A taxonomy of problem-based learning methods. *Medical Education, 20*(6), 481–486.

Becerra-Labra, C., Gras-Martí, A., & Martínez-Torregrosa, J. (2012). Effects of a problem-based structure of physics contents on conceptual learning and the ability to solve problems. *International Journal of Science Education, 34*(8), 1235–1253. https://doi.org/10.1080/09500693.2011.619210

Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology, 15*(3), ep430. https://doi.org/10.30935/cedtech/13176

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Díaz-Santamaría, J. O., Abril-Neira, A. M., Díaz-Santamaría, L. S., & Bravo-Faytong, F. A. (2017). The use of project-based learning in the teaching of kinematics. *Revista Atlante: Cuadernos de Educación y Desarrollo, 2*(7), 1–15.

Kanyesigye, S. T., Uwamahoro, J., & Kemeza, I. (2022). Difficulties in understanding mechanical waves: Remediated by problem-based instruction. *Physical Review Physics Education Research, 18*(1), 010140. https://doi.org/10.1103/PhysRevPhysEducRes.18.010140

Klein, P., Becker, S., Küchemann, S., & Kuhn, J. (2021). Test of understanding graphs in kinematics: Item objectives confirmed by clustering eye movement transitions. *Physical Review Physics Education Research, 17*(1), 013102. https://doi.org/10.1103/PhysRevPhysEducRes.17.013102

Liang, Y., Zou, D., Xie, H., & Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learning Environments, 10*, Article 52. https://doi.org/10.1186/s40561-023-00273-7

Marcinauskas, L., Iljinas, A., Čyvienė, J., & Stankus, V. (2024). Problem-based learning versus traditional learning in physics education for engineering program students. *Education Sciences, 14*(2), 154. https://doi.org/10.3390/educsci14020154

Morales, P., & Landa, V. (2004). Problem-based learning. *Theoria, 13*, 145–157.

Muñoz-Alvarez, G., Greca, I. M., & Arriassecq, I. (2025). Problem-based learning as a strategy for teaching physics in technical–professional higher education: A case study in Chile. *Education Sciences, 15*(8), 941. https://doi.org/10.3390/educsci15080941

Nicholus, G., Muwonge, C. M., & Joseph, N. (2023). The role of problem-based learning approach in teaching and learning physics: A systematic literature review. *F1000Research, 12*, 951. https://doi.org/10.12688/f1000research.136339.2

Polverini, G., & Gregorcic, B. (2024). Performance of ChatGPT on the test of understanding graphs in kinematics. *Physical Review Physics Education Research, 20*(1), 010109. https://doi.org/10.1103/PhysRevPhysEducRes.20.010109

Copyright: 2026 By the authors. This article is open access and distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license https://creativecommons.org/licenses/by-nc-sa/4.0/