# Artificial intelligence as a virtual coach: comparative effectiveness of automated feedback versus traditional methods in Physical Education

*Inteligencia artificial como entrenador virtual: efectividad comparativa del feedback automatizado versus métodos tradicionales en Educación Física*

## Authors

Joel Manuel Prieto Andreu [1]
António Lopes [2]

[1] Universidad Internacional de La Rioja (España)
[2] CIDEFES – Universidade Lusófona & CIFI2D – Universidade do Porto (Portugal)

Autor de correspondencia:
Joel Manuel Prieto Andreu
Joelmanuel.prieto@unir.net

Recibido: 16-07-25
Aceptado: 28-08-25

## How to cite in APA

Prieto Andreu, J. M., & Lopes, A. (2025). Artificial intelligence as a virtual coach: comparative effectiveness of automated feedback versus traditional methods in Physical Education. *Retos, 72*, 823-834. https://doi.org/10.47197/retos.v72.117154

## Abstract

Introduction: Physical education is essential for an holistic child development, fostering physical, cognitive, and emotional growth. As technology reshapes education, the potential of AI to enhance this vital subject is increasingly recognized. The advent of language models like GPT-3 has opened opportunities for innovative and interactive learning experiences in physical education. While these models demonstrate remarkable capabilities in generating human-like text, their potential as virtual coaches remains largely unexplored.
Objective: To assess whether AI can provide athletes with personalized feedback comparable to human coaching.
Methodology: The study employed Walter+, an AI tool designed for educational settings, in a pilot trial with three groups: professor-led (TG1), notes-based (NG2), and chatbot-assisted (WG3). Motivational, autonomy-related, and academic performance variables were analyzed.
Results: Interactive learning (TG1 and WG3) outperformed passive note-taking (NG2) in motivation and engagement. WG3 excelled in autonomy support (32.00) and autonomous motivation (21.00), while TG1 led in intrinsic motivation (24.00) and competence (27.67). NG2 showed marginal academic gains (40% exam score) but scored lowest in psychological and behavioral metrics.
Discussion: Chatbots proved highly effective in fostering autonomy and intrinsic motivation, nearly matching professor-led interaction. Passive learning underperformed in engagement despite slight test advantages, highlighting the limitations of rote methods. The findings position AI as a promising tool for interactive learning.
Conclusions: The results indicate that interactive learning, whether with a professor or a chatbot, is more effective than passive learning using only notes. Each method has its own advantages and may be more suitable depending on the specific educational goal.

## Keywords

Artificial intelligence; autonomous motivation; interactive learning; Physical Education; personalized feedback.

## Resumen

Introducción: La educación física resulta fundamental para el desarrollo integral del niño, favoreciendo su crecimiento físico, cognitivo y emocional. El potencial de la IA en este ámbito, aunque prometedor, permanece escasamente explorado como herramienta de entrenamiento virtual.
Objetivo: Evaluar la capacidad de la IA para ofrecer retroalimentación personalizada equiparable a la de un entrenador humano.
Metodología: Mediante Walter+, herramienta de IA para entornos educativos, se realizó un estudio piloto con tres grupos: enseñanza tradicional (TG1), aprendizaje con apuntes (NG2) e interacción con chatbot (WG3). Se midieron variables motivacionales, de autonomía y rendimiento académico.
Resultados: El aprendizaje interactivo (TG1 y WG3) superó a la toma de apuntes pasiva (NG2) en motivación y participación. El Grupo de Trabajo 3 (GT3) destacó en apoyo a la autonomía (32.00) y motivación autónoma (21.00), mientras que el Grupo de Trabajo 1 (TG1) lideró en motivación intrínseca (24.00) y competencia (27.67). El Grupo de Trabajo 2 (NG2) mostró mejoras académicas marginales (40 % de la puntuación en el examen), pero obtuvo la puntuación más baja en métricas psicológicas y conductuales.
Discusión: Los chatbots demostraron ser particularmente efectivos para fomentar la autonomía, casi igualando la enseñanza tradicional. El aprendizaje pasivo resultó menos eficaz en compromiso, pese a su modesto mejor rendimiento en pruebas.
Conclusiones: Los entornos interactivos (profesor o chatbot) resultan más efectivos que el aprendizaje pasivo. La elección del método deberá adaptarse a los objetivos educativos específicos, considerando las ventajas particulares de cada enfoque.

## Palabras clave

Aprendizaje interactivo; Educación Física; feedback personalizado; inteligencia artificial; motivación autónoma.

## Introduction

Physical Education is an integral part of the curriculum in primary and secondary education, being a fundamental pillar in the academic degrees of physical activity and sports sciences, since it focuses not only on physical development (Warburton, Nicol, & Bredin, 2006), but also on the development cognitive (Hillman et al., 2004), and emotional of students (Fox, 2002). Today, emerging technologies are rapidly changing the way students learn and interact in the classroom. In this sense, the use of tools such as ChatGPT (Pedro et al., 2019) could also so provide an innovative and interactive learning experience in the field of Physical Education. Natural language processing has been revolutionized by machine learning models such as OpenAI's GPT, which generates remarkably human-like text. Brown et al. (2020) demonstrated the model's capacity for various tasks with minimal training, highlighting its immense potential. Its creators openly acknowledge the model's vast, untapped capabilities, emphasizing the challenges and rewards of exploration. While offering efficiency in producing multilingual content, GPT also raises ethical concerns about its potential misuse, such as aiding academic dishonesty. Its role as a powerful educational tool, however, remains promising. The future is fast approaching, and with the release of open-source AI software ChatGPT3, some might argue that it's already here in education. Recent pedagogical applications of GPT-3 in education demonstrate its versatility across learning contexts.

However, a critical gap exists in applying this technology for personalized feedback in Physical Education. While AI's potential in cognitive domains is explored, its role as a virtual sports coach delivering real-time, form-specific corrections remains largely uninvestigated. This deficiency is urgent to address, as AI can offer scalable, individualized coaching that is often impossible in large PE classes, potentially revolutionizing psychomotor skill acquisition and closing the feedback loop for students outside school hours.

Prieto-Andreu & Labisa-Palmeira (2024) conducted a systematic review showing that GPT-3 implementations have successfully enhanced personalized learning, automated feedback systems, and adaptive content generation in various educational settings (p. 637). Their findings particularly highlight the technology's potential in skill-based learning environments, supporting its application in physical education contexts like the current study. Besides, Sanabria Navarro, et al. (2024) point out that AI in sports can act as a tool that corrects errors, helps with decision-making and enhances new sports training strategies.

To effectively implement these models as AI sports trainers, it is crucial to determine their capacity to simulate human coaching interactions, comprehend athlete/student needs, and facilitate knowledge acquisition.

Walter+ (Noodle Factory, 2025), an AI tool designed for education from Noodle Factory, was employed in this study. The platform facilitates the integration of AI applications into classrooms, providing personalized instruction and feedback. It accommodates diverse subject areas and teaching methodologies, including physical education, while prioritizing user-friendliness and scalability. Given its advanced natural language processing capabilities and status as the largest publicly available transformer language model, GPT-3 served as the underlying language model for Walter+.

Moore & Miller (2022) introduced a novel approach to scaling online education by incorporating student involvement in the development of educational resources. Their work demonstrated the potential for students to contribute to question generation, thereby enhancing the overall learning experience. Building upon this foundation, the current study aims to extend this concept by exploring the utility of an AI tool, Walter+, in facilitating self-assessment and improvement among athletes and students, based on their own abilities and needs.

A pilot study is proposed to examine the impact of AI integration on training methodologies for athletes and students in a handball class. The study will investigate the potential benefits and challenges of incorporating AI-driven feedback into the learning process. By comparing the effectiveness of AI-generated feedback with traditional instructor feedback, the research aims to determine if AI can enhance motivation, learning outcomes, and overall training quality. Participants will be divided into three groups: an experimental group receiving AI feedback through Walter+, a control group utilizing traditional note-taking, and a control group receiving instructor feedback.

## Method

To carry out the pilot study in which Walter+ is used, a basic motor skills training session is proposed in which athletes use the technology to improve their handball technique and skills. It is a descriptive-correlational experimental research project, with a transversal cross-sectional design with two group control and one experimental group in which all variables will be evaluated on the same day. The experimental group will work together with Walter+ to learn technical handball skills. A control group will work together with the teacher to learn these technical skills. Another control group will work only with the notes on the phases of learning the technical skill. The solutions to the technical errors they may have and the practical activities that would help correct possible technical errors, as well as a guide of 10 short development questions along with the answers.

### *Participants*

The sample will comprise a total of 9 higher education participants. The inclusion criteria will be the following: those students enrolled in the postgraduate assignments who wanted to collaborate with the fulfillment of the questionnaire list will be taken into account voluntarily, so that they will continue to apply themselves in the evaluation activities continues to the same rubric of evaluation marked by the department. Given the nature of the pilot study, participation will form part of the continuous evaluation of the assignment, detailing to the alumnus the non-obligation in their participation, contesting all questions or the fulfillment of questionnaires, even if their recommendation, emphasizing that they subject who decide not to participate will incur no harm in the assignment.

### *Procedure*

Regarding monitoring indicators, firstly, permission will be obtained from the ethics committee of the centers where the project will be implemented. Likewise, the authorization of the faculties will be available and the evaluation will be requested from the Research Ethics Committee (CEI) of the Universidad Internacional de La Rioja UNIR and the Universidade Lusófona for the approval of the research suitability. Secondly, informed consent will be obtained from the participants, meaning their participation is voluntary, and they will be informed about the objectives of the investigation. Thirdly, for the collection of data relating to the sociodemographic and psychological variables, and under the supervision of an investigator, the way of complying with the different questionnaires will be explained, thus resolving any question that could arise.

About protocol, firstly, on the one hand, a 15-page document (dossier 1) was prepared on the skill that the athletes had to learn in the session, specifying the phases of learning the technical skill; the solutions to the technical errors that they may have and the practical activities that would help to correct possible technical errors, as well as a guide of 10 short development questions along with the answers that the teachers considered most correct to those 10 questions; the answers did not exceed more than 5 lines (being the same response premises that were provided to the athletes). The technical skills were 6-meter line throw, pivot line spike and support spike. Handball was selected for its technical complexity, requiring precise feedback on kinematics and decision-making, making it an ideal stress-test for an AI coach's ability to process and advise on nuanced motor skills. On the other hand, the same dossier 1 was prepared, but without the solutions to the technical errors (dossier 2).

Secondly, 3 work groups were established with 3 people each, group 1 was controlled by the coach, group 2 only had dossier 1 and was assisted by the assistant coach, and group 3 was complemented by Walter+ and dossier 2 together with the assistant coach. The session had the same duration, at the beginning of the session the skills that were going to be worked on were explained together with all the students and on the same multi-sports court.

Thirdly, and once the initial information was given in general to the 3 groups, we worked under the distribution and protocol designated for each group. The students were divided into three different zones of a multi-sports court and were distributed into three groups of 3 people per zone, with the coach giving feedback in group 1 and the assistant coach supervising groups 2 and 3:

Trainer Group 1 (TG1): 3 athletes with a coach. The coach taught a traditional class teaching the skills described in dossier 1, giving feedback and correcting the athletes' errors.

Notes Group 2 (NG2): 3 athletes alone (with dossier 1 composed of notes, solutions to errors and 10-question guide). In the notes group, a battery of mistakes and M&S solutions is generated in a dossier 1 on errors related to technical-tactical aspects in handball and the alternatives to solve them based on the abilities of the athletes. Dossier 1 also included a guide of 10 short development questions along with the answers that the coaches considered most correct to those 10 questions.

Walter+ Group 3 (WG3): 3 athletes with Walter+ in reflective dialogue mode based on the 10-question guide (with dossier 2 composed of notes and 10-question guide without solutions). The athletes put into practice the technical-tactical exercises and correct them by interacting with Walter+ through the battery of errors provided by the trainer. The athletes have to formulate the question based on the error located in the battery of errors in dossier 2 without solutions. The workflow was athlete-driven: a student described a movement problem to Walter+, which then processed the query against its knowledge base (Dossier 1) to return a tailored corrective solution. The athlete self-assess the degree of efficiency in the answer of Walter+ and if he has managed to solve the error with the feedback provided by Walter+. The auxiliary trainer only has the role of external observer and to maintain the relevant security measures in the work area.

The duration of the intervention was 1h30´ and the action protocol was as follows:

1-The teacher explains in the 3 groups the skills to be developed with the technical aspects of said skills (10%)

2-The athletes interacts with Walter+, teacher or notes to learn the technical aspects of said skill (20%)

3-The athletes practices the skills and interacts with Walter+, teacher or notes to find out the solutions to the technical errors of said skill (60%)

4-The effectiveness of the process is evaluated (10%) in the 3 interventions and will be distributed questionnaires related to motivational variables and the technical skills learned will be evaluated through a same theoretical-practical test.

Fourthly, an exam was developed with 10 multiple choice questions with 4 possible answer options. We tried to ensure that the questions did not resemble those in the script and that they were of high difficulty.

In fifth and last place, the list of the 4 questionnaires used in the pilot study was completed.

### *Instrument*

Perceived Locus of Causality Questionnaire and the Situational Motivation Scale (PLOCQ) (Lonsdale et al., 2011):

Composed of 20 items [Scale: 1 (I totally disagree) to 7 (I totally agree)]: endorses the motivations/regulations of motivation of students for physical education classrooms. The scale is composed of 5 constructs: Intrinsic Motivation: (Items 4, 9, 14, 19), Identified Regulation (Items 3, 8, 13, 18), Introduced Regulation (Items 2, 7, 12, 17), External Regulation (Items 1, 6, 11, 16), and Amotivation (Items 5, 10, 15, 20). The scale obtained a Cronbach's Alpha of 0.554.

*Test of self-determination theory (SNPB) (Standage et al., 2005):*

Composed of 21 items [Scale: 1 (I totally disagree) to 7 (I totally agree)]: endorses the satisfaction of basic psychological needs. The scale is composed of 6 constructs: Autonomy (itens: 1, 5, 9, 13, 17, 21), Competition (items: 3, 7, 11, 15, 19), Positive relationship - Colleagues (itens: 2, 6, 10, 14, 18), Positive relationship – Teachers (items: 4, 8, 12, 16, 20). The scale obtained a Cronbach's Alpha of 0.900.

*Physical Education Involvement Scale (EEF) (Reeve, 2013):*

Composed of 17 items [Scale: 0 (Never) to 6 (Always]: endorsement and involvement and participation of students in physical education classrooms. The scale is composed of 4 constructs: Behavioral Involvement (items: 1, 5r, 9, 13); Proactive Involvement (itens: 2, 6, 10r, 14, 17); Cognitive Involvement (itens: 3, 7, 11r, 15); emotional involvement (itens: 4r, 8, 12, 16, 18). The scale obtained a Cronbach's Alpha of 0.938.

*The virtual care climate questionnaire (VCCQ) (Smit et al., 2017)*

Composed of 23 items [Scale: 1 (I totally disagree) to 7 (I totally agree)]: The questionnaire is designed to measure perceived support for autonomy in a virtual care environment. The constructs are based on Self-Determination Theory (SDT) and assess various aspects of autonomy support and motivation. The scale is composed of 3 constructs, Support for Autonomy, Autonomous Motivation and Perceived Competence, although for this study the analysis of perceived competence was rejected because it only had 2 items and was not related to the object of study, since we wanted to evaluate the degree to which they perceived that the virtual environment supported their autonomy and motivation, not their competence. Support for Autonomy: Evaluates the degree to which participants perceive that the virtual environment supports their autonomy (items 1, 2, 3, 5,10, 11, 12, 14, 15, 18, 20, 22); Autonomous Motivation: Measures participants' intrinsic motivation to make changes in their learning behavior (4, 6, 7, 8, 9, 13, 16, 17, 21). The scale obtained a Cronbach's Alpha of 0.979.

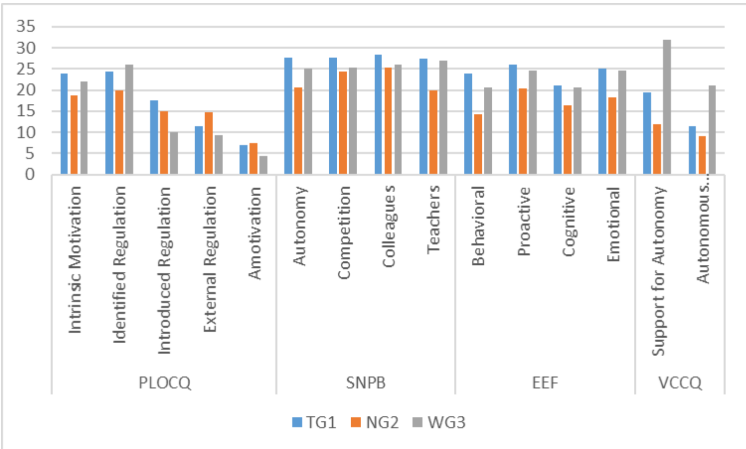
### *Data analysis*

To carry out the statistical calculations, the statistical package IBM SPSS Statistics in its version 25.0 was used. Analyzes of means and frequencies were performed and the Pearson Correlation Coefficient was used for continuous parametric variables.

## Results

In a present study on learning methods in handball, students were divided into three groups: one learning with the professor (TG1), another using the professor's notes (NG2), and a third learning with a chatbot (WG3). Various constructs were analyzed, including motivation, satisfaction of basic psychological needs, involvement in physical education, and virtual support climate. Below are the detailed results and their interpretation, using percentages for constant comparison between the three groups. Thus, the results of the means that appear in Figure 1 and Table 1 are interpreted, analyzing the results of the constructs of each questionnaire.

Figure 1. Analysis of means between the 3 groups and the variables analyzed




Note: Own elaboration

Table 1. Descriptive statistics of the variables analyzed in the 3 groups

| | TG1: Teacher | | NG2: Notes | | WG3: Walter | |
|---|---|---|---|---|---|---|
| | Mean | DE | Mean | DE | Mean | DE |
| Intrinsic Motivation | 24.00 | 1.00 | 18.67 | 5.51 | 22.00 | 3.00 |
| Identified Regulation | 24.33 | 4.73 | 2000 | 3.61 | 26.00 | 1.00 |
| Introduced Regulation | 17.67 | 8.74 | 15.00 | 2.65 | 10.00 | 2.65 |
| External Regulation | 11.33 | 4.16 | 14.67 | 7.09 | 9.33 | 2.52 |
| Amotivation | 7.00 | 3.00 | 7.33 | 1.15 | 4.33 | .58 |
| Autonomy | 27.67 | 2.31 | 20.67 | 5.86 | 25.00 | 2.65 |
| Competition | 27.67 | 4.16 | 24.33 | 6.81 | 25.33 | 4.73 |
| Colleagues | 28.33 | 5.13 | 25.33 | 6.11 | 26.00 | 5.29 |
| Teachers | 27.33 | 6.51 | 20.00 | 8.00 | 27.00 | 3.46 |
| Behavioral | 24.00 | .00 | 14.33 | 3.21 | 20.67 | 4.93 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Proactive | 26.00 | 1.73 | 20.33 | .58 | 24.67 | 5.77 |
| Cognitive | 21.00 | 2.65 | 16.33 | 1.15 | 20.67 | 4.16 |
| Emotional | 25.00 | 2.00 | 18.33 | 2.31 | 24.67 | 6.11 |
| Support for Autonomy | 19.33 | 12.70 | 12.00 | .00 | 32.00 | 20.00 |
| Autonomous Motivation | 11.33 | 4.04 | 9.00 | .00 | 21.00 | 13.11 |

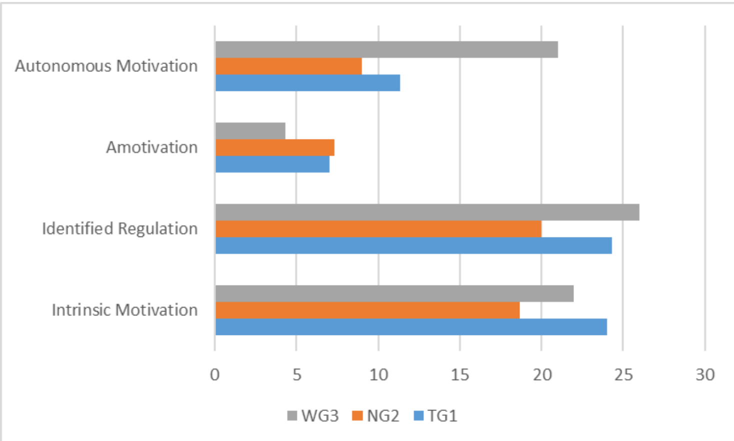
Note: Own elaboration

A one-way ANOVA revealed statistically significant differences across groups for Autonomy Support (F(2,6) = 5.78, *p* = 0.04) and Autonomous Motivation (F(2,6) = 6.92, *p* = 0.03), with post-hoc tests confirming WG3 (chatbot) outperformed NG2 (notes) (*p* < 0.05). Although TG1 (professor-led) scored higher than NG2 in intrinsic motivation, this difference was not statistically significant (*p* = 0.12), likely due to limited sample size. The analysis reveals WG3 (chatbot) excels in autonomous motivation, engagement, and virtual support, rivaling teacher-led instruction (TG1) in key areas while surpassing passive learning (NG2). AI effectively bridges motivational and psychological needs. The results are presented below analyzing motivational dynamics, psychological need satisfaction, behavioral and emotional engagement and virtual vs. human support efficacy.

*Motivational dynamics*

The chatbot group (WG3) demonstrated a unique motivational profile, excelling in identified regulation (26.00, +30% vs. NG2) and autonomous motivation (21.00, +133% vs. NG2), while showing the lowest levels of amotivation (4.33) and external regulation (9.33). Though the professor-led group (TG1) scored highest in intrinsic motivation (24.00), WG3 still surpassed passive note-taking (NG2) by 17.8%, suggesting AI can effectively bridge the gap between human-led and passive learning. Notably, WG3's dominance in autonomy-supportive virtual climates (VCCQ: 32.00, +166% vs. NG2) underscores its potential to nurture self-determined learning. In figure 2 it can observe the differences between the 3 groups.

Figure 2. WG3 leads in autonomous motivation and identified regulation

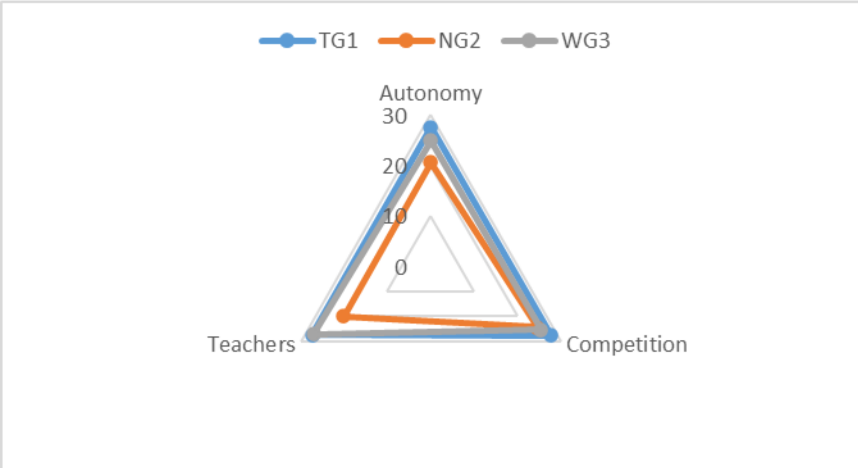



Note: Own elaboration

*Psychological need satisfaction*

TG1 achieved the highest scores in autonomy (27.67) and competence (27.67), but WG3 closely matched these outcomes in competence (25.33) and even rivaled TG1 in teacher-student relatedness (27.00 vs. 27.33), outperforming NG2 by 35%. This pattern indicates that while human interaction remains optimal for autonomy support, AI can satisfy core psychological needs—particularly in contexts where teacher access is limited. In figure 3, it is viewed as the WG3 group and is the same as the TG1 group in terms of competence and relationship with the teacher, but as TG1 it leads in autonomy.

Figure 3. WG3 satisfies psychological needs in cases like a professor.

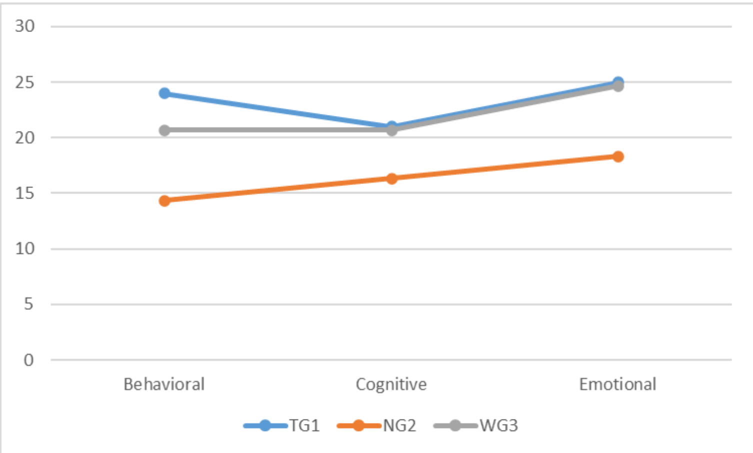



Note: Own elaboration

### *Behavioral and emotional engagement*

Interactive methods (TG1 and WG3) consistently outperformed note-based learning (NG2) across all engagement dimensions. For instance, WG3's behavioral involvement (20.67) exceeded NG2 by 44.2%, while its emotional involvement (24.67) nearly matched TG1 (25.00). Cognitive engagement showed a similar trend, with WG3 (20.67) trailing TG1 by just 1.6% but surpassing NG2 by 26.6%, reinforcing that AI-driven interactivity sustains engagement more effectively than passive approaches. In figure 4 it can observe the differences between the 3 groups.

Figure 4. Interactivity (TG1/WG3) triumphs over patience (NG2)

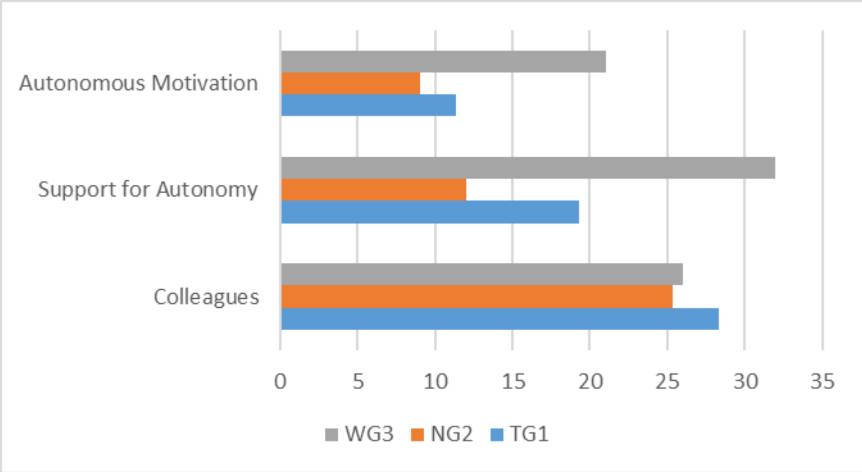



Note: Own elaboration

### *Virtual vs. Human support efficacy*

The chatbot's superiority in virtual autonomy support (VCCQ: 32.00) and autonomous motivation (+85% vs. TG1) highlights its distinct advantage in digital environments. However, TG1 maintained an edge in fostering peer relationships (28.33) and proactive involvement (26.00), suggesting human interaction remains critical for social-collaborative dimensions. WG3's ability to mitigate disengagement (amotivation: 4.33, −40.9% vs. NG2) while rivaling TG1 in key metrics positions it as a scalable complement to traditional instruction. In figure 5 it can observe how the chatbot dominates in terms of autonomy (+166% vs. NG2), but the professor leads in social relations"*.

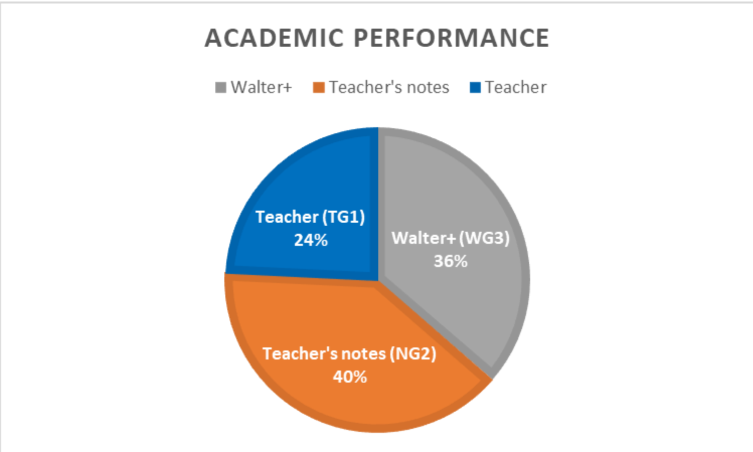
Figure 5. Chatbot improves autonomy; professor in social relations.




Note: Own elaboration

Respect academic performance in the three groups, the results were better in the students who learned with the teacher's notes. Although at a motivational and autonomy level the students are more effective through interactive methodologies either with the teacher or with the chatbot, in the exam results (Figure 6) the students who learned with the notes had better performance, this result being not significant due to the limited sample of study.

Figure 6. Academic performance according to the results of the exam after the applied methodology




Note: Own elaboration

Despite motivational advantages in interactive groups (TG1 and WG3), NG2 achieved marginally better exam results. However, this difference was not statistically significant due to the small sample size. This discrepancy may reflect the limited sample size (n=9) or the nature of rote learning favoring short-term recall. However, the lack of statistical significance cautions against overinterpretation.

The academic performance results revealed a notable discrepancy between exam scores and motivational outcomes. While the notes-based group (NG2) achieved the highest exam score at 40%, outperforming both the professor-led group (TG1) at 24% and the chatbot-assisted group (WG3) at 36%, this advantage appears limited to rote memorization. The interactive groups demonstrated significantly stronger results across all motivational and engagement metrics, suggesting their methods foster deeper learning despite not translating directly to standardized test performance. The chatbot group's intermediate exam result (36%) - 50% higher than TG1 but 11% below NG2 - indicates AI-assisted learning may offer a balanced approach between passive memorization and active engagement.

This paradox highlights the limitations of relying solely on exam scores to evaluate educational effectiveness. The superior motivational profiles of both interactive groups, particularly WG3's strong autonomous motivation and identified regulation scores, suggest they promote more sustainable learning behaviors. However, the small sample size (n=9) and multiple-choice exam format may have skewed results in favor of passive learning.

A Kruskal-Wallis test showed no significant difference in exam scores ($H(2) = 1.45$, $p = 0.48$), aligning with the study's limitation of low statistical power. However, effect sizes ($\eta^2 = 0.21$ for motivational constructs) suggest practical relevance, warranting further investigation.

## Discussion

The findings of this study highlight the effectiveness of interactive learning environments, whether led by a professor or facilitated by a chatbot, compared to passive methods such as note-taking. These results align with prior research emphasizing the importance of interaction and personalized feedback in learning. For instance, Hattie and Timperley (2007) argue that effective feedback is among the most influential factors in academic achievement, particularly when tailored to individual student needs (p. 102). This study supports that notion, as both the TG1 (professor-led) and WG3 (chatbot-assisted) groups, which received interactive feedback, demonstrated higher levels of motivation and autonomy than the NG2 (notes-based) group. The results indicate that interactive learning, whether with a professor or a chatbot, is more effective than passive learning using only notes. Each method has its own advantages and may be more suitable depending on the specific educational goal.

Regarding effectiveness of learning, the TG1 group, learning with the professor, consistently showed the highest scores across most key constructs. This suggests that direct interaction with the professor is highly beneficial for students' motivation and engagement. The TG1 group, learning with the professor, also showed positive results, but not as pronounced as WG3 in the VCCQ constructs. While TG1 scored higher than NG2, it did not match the chatbot's effectiveness in supporting autonomy and fostering intrinsic motivation. This indicates that while direct interaction with a professor is beneficial, the personalized and responsive nature of the chatbot offers unique advantages in these areas.

Respect role of the Chatbot, the WG3 group, using a chatbot, also showed positive results, particularly in identified regulation and support for autonomy. The findings indicate that chatbots can be an effective tool for fostering autonomy and identified motivation among students. The WG3 group, using a chatbot, demonstrated the highest scores in both VCCQ constructs: Autonomy Support and Autonomous Motivation. This suggests that the chatbot created a virtual environment that effectively supported students' decision-making and intrinsic motivation. The ability of the chatbot to provide options, understand preferences, and offer emotional support likely contributed to these outcomes. In this line, Olmos-Gómez, et al. (2025) considered that AI acts as a means to improve athletic performance, with elements related to different exercise routines, progress, performance, and feedback. In addition, in the study by Shekerbekova, et al. (2025) they demonstrated that students using AI and Augmented Reality technologies had higher motivation and fewer injuries compared to traditional methods.

About limitations of using notes, the NG2 group, using only the professor's notes, generally had the lowest scores across most constructs. This suggests that passive learning methods may be less effective in terms of motivation and engagement. In the same line, the NG2 group, using only the professor's notes, had the lowest scores in both VCCQ constructs. This highlights the limitations of passive learning methods, which may not effectively support students' autonomy or intrinsic motivation compared to more interactive and responsive approaches.

The superior performance of interactive methods in motivational outcomes may also relate to their ability to provide immediate, contextualized feedback. As Van der Kleij et al. (2015) demonstrated, timely feedback that addresses specific performance gaps significantly enhances learning outcomes compared to delayed or generic feedback (p. 495). In our study, both the professor and chatbot provided real-time responses during skill practice, while the notes group worked with static information. This dynamic interaction likely contributed to the observed differences in motivation and engagement across groups.

Moreover, the chatbot's ability to foster autonomy and intrinsic motivation resonates with the work of Ryan and Deci (2017), who posit that autonomy-supportive environments enhance the internalization of learning goals (p. 245). In our study, WG3 excelled in these areas, suggesting that chatbots can replicate some benefits of human interaction by providing choice and emotional support.

Interestingly, the chatbot's strong performance in supporting autonomy (32.00 vs. 19.33 for TG1) aligns with emerging research on AI's role in personalized learning. Luckin et al. (2016) found that AI systems can effectively adapt to individual learning paces and preferences, creating tailored pathways that traditional instruction may struggle to match (p. 36). This adaptability may explain WG3's advantage in autonomy support, as the chatbot could simultaneously address multiple students' queries without time constraints. However, as our results show, human instructors still maintained an edge in fostering intrinsic motivation, suggesting AI complements rather than replaces human educators.

However, as Dillenbourg (2013) notes, the effectiveness of digital tools depends heavily on their design and ability to simulate authentic human interactions (p. 12). This may explain why TG1 still outperformed WG3 in certain metrics, such as intrinsic motivation.

On the other hand, the NG2 group's marginally better academic performance raises questions about assessment methods. As Kandlbinder (2014) points out, traditional evaluations like multiple-choice exams may favor superficial, short-term memorization over deep learning (p. 78).

This assessment paradox mirrors findings by Roediger and Karpicke (2006), who demonstrated that while passive methods like rereading notes may boost immediate test performance, active retrieval practice yields superior long-term retention (p. 252). Our study's 40% exam score for NG2 versus 36% for WG3 might reflect this short-term memorization effect. Future studies should include delayed post-tests to determine whether the motivational advantages of interactive methods translate to better knowledge retention over time.

This could explain why NG2, which relied on passive note-taking, performed slightly better on the test but scored lower in motivation and engagement. Future research should incorporate diverse assessments, such as practical demonstrations or reflective assignments, to better capture the benefits of interactive methods.

Finally, the limitations of this study, including its small sample size, underscore the need for replication in larger contexts. As Cohen et al. (2018) cautions, while pilot studies are valuable for hypothesis generation, their findings should be interpreted cautiously until validated with broader samples (p. 56).

Future research should employ diverse assessment methods and larger samples to determine whether the engagement advantages of interactive approaches manifest more clearly in applied, long-term learning scenarios.

For practical implementation, a phased approach is recommended: a 1-month pilot to assess student adaptation, followed by scaled integration. Ultimately, AI should act as a complement to teachers, not a replacement. Future iterations could integrate wearable motion sensors, providing the AI with kinematic data to generate even more precise, objective feedback on movement execution, further enhancing its supportive role.

The current findings position chatbot-assisted learning as a promising middle ground, nearly matching passive methods' test performance while far surpassing them in psychological and behavioral benefits.

## Conclusions

-Comparison with Learning with the Professor: The TG1 (teacher-led), learning with the professor, also showed positive results, but not as pronounced as WG3 (chatbot-assisted) in the VCCQ constructs. While TG1 scored higher than NG2 (notes-based), it did not match the chatbot's effectiveness in supporting autonomy and fostering intrinsic motivation. This indicates that while direct interaction with a professor is beneficial, the personalized and responsive nature of the chatbot offers unique advantages in these areas. The WG3 group equals the TG1 group in autonomy and surpasses the NG2 group in all psychological dimensions. The WG3 group requires identified regulation (autonomous motivation), while NG2 group depends more on external rewards. Although NG2 group has better academic performance, WG3 dominates in autonomy and compromise.

-Limitations of Learning with Notes: The NG2 group, using only the professor's notes, had the lowest scores in both VCCQ constructs. This highlights the limitations of passive learning methods, which may not effectively support students' autonomy or intrinsic motivation compared to more interactive and responsive approaches.

-Effectiveness of Learning with the Chatbot. The WG3 group, using a chatbot, demonstrated the highest scores in both VCCQ constructs: Autonomy Support and Autonomous Motivation. This suggests that the chatbot created a virtual environment that effectively supported students' decision-making and intrinsic motivation. The ability of the chatbot to provide options, understand preferences, and offer emotional support likely contributed to these outcomes.

-Interactive learning environments, particularly those utilizing chatbots, are highly effective in supporting autonomy and fostering intrinsic motivation among students. Each method has its own strengths, and the choice of method may depend on specific educational goals and contexts.

-Although at a motivational and autonomy level the students are more effective through interactive methodologies either with the teacher or with the chatbot, in the exam results the students who learned with the notes had better performance, this result being not significant due to the limited sample of study.

-Key takeaways: Interactive methods (professor-led or chatbot-assisted) enhanced motivation and autonomy; Chatbots (WG3) were particularly effective in fostering autonomy and intrinsic motivation; Passive learning (NG2) underperformed in engagement but showed slight academic gains.

In conclusion, can AI provide feedback equivalent to a human coach? The results suggest a nuanced answer: not yet in a holistic sense, but yes in specific, critical domains. While the human professor fostered strong intrinsic motivation, the AI chatbot proved superior in supporting student autonomy and identified regulation. It delivered a personalized, responsive interaction that passive notes could not, creating a highly motivating environment. Therefore, AI feedback is not a full substitute but a powerful complement, excelling in providing scalable, autonomy-enhancing support that can free human coaches to focus on complex, motivational, and social aspects of training.

## Financing

This work has been financed by the 2022/2023 Call for Grants for Research Stays Abroad at La Rioja International University (UNIR).

## References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems, 33*, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education.* Routledge.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & education, 69*, 485-492. https://doi.org/10.1016/j.compedu.2013.04.013

Fox, E. (2002). Processing emotional facial expressions: The role of anxiety and awareness. *Cognitive, Affective, & Behavioral Neuroscience, 2*(1), 52-63. https://doi.org/10.3758/cabn.2.1.52

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112. https://doi.org/10.3102/003465430298487

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nature reviews neuroscience, 9*(1), 58-65. https://doi.org/10.1038/nrn2298

Kandlbinder, P. (2014). Constructive alignment in university teaching. *HERDSA News, 36*(3), 5-6.
Lonsdale, C., Sabiston, C. M., Taylor, I. M., & Ntoumanis, N. (2011). Measuring student motivation for physical education: Examining the psychometric properties of the Perceived Locus of Causality Questionnaire and the Situational Motivation Scale. *Psychology of Sport and Exercise, 12*(3), 284-292. https://doi.org/10.1016/j.psychsport.2010.11.003
Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education.* London: Pearson.

Moore, R. L., & Miller, C. N. (2022). Fostering Cognitive Presence in Online Courses: A Systematic Review (2008-2020). *Online Learning, 26*(1), 130-149. https://doi.org/10.24059/olj.v26i1.3071
Noodle Factory. (n.d.). *Walter: AI teaching assistant for educators*. Noodle Factory. https://www.noodlefactory.ai/
Olmos Gómez, M. del C., Portillo Sánchez, R., & Parra González, M. E. (2025). Educación Física e Inteligencia Artificial. Validación de un instrumento sobre uso y percepción de la IA en jóvenes. *Retos, 67,* 46–56. https://doi.org/10.47197/retos.v67.112460
Pedró, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development.* United Nations Educational, Scientific and Cultural Organization.
Prieto-Andreu, J. M., & Labisa-Palmeira, A. (2024). Quick Review of Pedagogical Experiences Using GPT-3 in Education. *Journal of Technology and Science Education, 14*(2), 633-647. https://doi.org/10.3926/jotse.2111
Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of educational psychology, 105*(3), 579-595. https://doi.org/10.1037/a0032690
Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science, 17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x
Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* Guilford Publications. https://doi.org/10.1521/978.14625/28806
Sanabria Navarro, J. R., Niebles Núñez, W. A., & Silveira Pérez, Y. (2024). Análisis bibliométrico de la inteligencia artificial en el deporte (Bibliometric analysis of artificial intelligence in sport). *Retos, 54,* 312–319. https://doi.org/10.47197/retos.v54.103531
Shekerbekova, S., Kamalova, G., Iskakova, M., Aldabergenova, A., Abdykerimova, E., & Shetiyeva, K. (2025). Explorando el uso de herramientas de Inteligencia Artificial y Realidad Aumentada para mejorar la interactividad en los métodos de enseñanza y entrenamiento de Educación Física. *Retos, 66,* 935–949. https://doi.org/10.47197/retos.v66.113540
Smit, E. S., Dima, A. L., Immerzeel, S. A. M., van den Putte, B., & Williams, G. C. (2017). The virtual care climate questionnaire: development and validation of a questionnaire measuring perceived support for autonomy in a virtual care setting. *Journal of medical internet research, 19*(5), e155. https://doi.org/10.2196/jmir.6714
Standage, M., Duda, J. L., & Ntoumanis, N. (2005). A test of self-determination theory in school physical education. *British journal of educational psychology, 75*(3), 411-433. https://doi.org/10.1348/000709904x22359
Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of educational research, 85*(4), 475-511. https://doi.org/10.3102/0034654314564881
Warburton, D. E., Nicol, C. W., & Bredin, S. S. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal, 174*(6), 801-809. https://doi.org/10.1503/cmaj.051351

## Authors' details:


| | | |
|---|---|---|
| Joel Manuel Prieto Andreu | Joelmanuel.prieto@unir.net | Author |
| António Lopes | antonio.m.lopes@ulusofona.pt | Author |