Articles

Artificial intelligence as a pedagogical tool for physical fitness and emotional well-being in basic education

Jorge Andrés Bayona Martínez [1]; Edwin Esneider Andrade Perdomo [2] & Steven Arturo Torres Burgos [3]

Jorge Andrés Bayona Martínez
Universidad Estatal de Milagro, Ecuador
jbayonam@unemi.edu.ec
ORCID https://orcid.org/0009-0005-9332-5764

Edwin Esneider Andrade Perdomo
Universidad estatal de Milagro, Ecuador
eandradep7@unemi.edu.ec
ORCID https://orcid.org/0009-0001-1200-9593

Steven Arturo Torres Burgos
Universidad estatal de Milagro, Ecuador
storresb5@unemi.edu.ec
ORCID https://orcid.org/0000-0001-9299-3254

RIAF. International Journal of Physical Activity
Universidad de Guayaquil, Ecuador
Frequency: Semi-annual
Vol. 4, N° 2, 2026
revista.riaf@ug.edu.ec

Received: June 5th, 2026
Approved: July 6th, 2026
Published: July 25th, 2026

URL: https://revistas.ug.edu.ec/index.php/riaf
DOI: https://doi.org/10.53591/riaf.v4i2.3372

Abstract: Introduction: Physical inactivity and emotional well-being difficulties are growing problems in the school population, whereas artificial-intelligence-based technologies offer new pedagogical possibilities. Objective: to assess the effect of a physical education program mediated by pedagogical possibilities. Objective: to assess the effect of a physical education program mediated by artificial intelligence on the physical fitness and emotional well-being of basic-education students. Materials and methods: a quasi-experimental design with two groups and pre- and post-measurements was used over twelve weeks. Two hundred and forty primary and secondary students participated and were assigned to an experimental group and a control group. Cardiorespiratory endurance was assessed with a twelve-minute run test, muscular endurance with elbow push-ups in thirty seconds, speed over fifty meters, and an emotional well-being index. Analyses of covariance, mixed analyses of variance, effect sizes with confidence intervals and power analysis were applied. Results: the experimental group showed significantly greater improvements than the control group in all variables, with significant group-by-time interactions and moderate-to-large effect sizes. Gains in physical fitness were positively associated with improvements in emotional well-being. Conclusions: the results, derived from simulated data for methodological purposes, suggest that artificial intelligence, used as a pedagogical tool, could simultaneously favor physical fitness and emotional well-being, although studies with empirical data are required to confirm these trends.

Keywords: Artificial intelligence; physical fitness; emotional well-being; basic education; educational technology.

Introduction

Regular physical activity constitutes a recognized determinant of health during childhood and adolescence. The World Health Organization recommends that children and adolescents aged five to seventeen accumulate at least sixty minutes of daily moderate-to-vigorous physical activity and incorporate muscle and bone-strengthening exercises at least three days per week (Bull et al., 2020). Despite these recommendations, most of the school-age population worldwide does not meet the minimum levels of physical activity, and the prevalence of inactivity tends to increase with age and is higher among girls (Marques et al., 2020). The most recent international reviews confirm that approximately eighty percent of adolescents are insufficiently active and that school, social, and digital environments are crucial for reversing this trend (van Sluijs et al., 2021). The global landscape of physical activity indicators reinforces this concern, evidencing persistently low grades in most countries (Aubert et al., 2022).

Physical fitness, understood as the ability to perform motor tasks efficiently, integrates components such as cardiorespiratory endurance, muscular strength-endurance, and speed, and has historically been linked to better health indicators and, to a lesser extent, to the academic performance of schoolchildren (Xu et al., 2023). Alongside the physical dimension, emotional well-being has gained relevance on the educational agenda, given that adolescence is a sensitive period for the emergence of symptoms of anxiety, stress, and low mood. Available evidence indicates that physical activity interventions produce small to moderate improvements in various mental health indicators of children and adolescents, with especially notable effects on stress and social competence (Fu et al., 2025). School-based controlled studies have documented reductions in depressive symptoms and increases in life satisfaction following multicomponent physical activity programs (Ahmed et al., 2023).

In this context, artificial intelligence has emerged as a set of technologies capable of personalizing teaching, automating learning assessment, and providing adaptive feedback. In the field of physical education, its usefulness has been proposed for individualizing classes, providing knowledge, evaluating students, and guiding accompaniment processes (Lee & Lee, 2021). Several recent developments show its application in teaching assessment, the design of motor simulation environments, and the improvement of the overall quality of school health (Li et al., 2024; Zhang et al., 2022). Complementarily, wearable devices with automated analytics have shown favorable, albeit heterogeneous, effects on the daily physical activity of the school population (Casado-Robles et al., 2022; Chen et al., 2025), and active video games have evidenced benefits on physical fitness, enjoyment, and psychological well-being in educational contexts (Marsigliante et al., 2024; Rosi et al., 2025).

In the socio-emotional dimension, digital technologies and artificial intelligence-supported approaches have been incorporated into social and emotional learning programs, whose efficacy on competencies, attitudes, and school climate has robust meta-analytic support (Cipriano et al., 2023). Likewise, the potential of conversational systems and emotional analytics applications to support the mental health of children and adolescents has been described, always under principles of human supervision, equity, and data protection (Dritsona et al., 2025). However, their adoption requires the development of competencies and artificial intelligence literacy in both students and teachers, as well as the consideration of ethical implications (Farrelly & Baker, 2023; Ng et al., 2024).

Despite growing interest, a gap persists in the state of knowledge: most of the literature addresses separately the effects of artificial intelligence on motor learning or on emotional well-being, without integrating both dimensions into a single design oriented towards basic education. Furthermore, available studies are concentrated in higher education and high-income contexts, and rarely articulate physical fitness tests with emotional well-being measures within a single pedagogical proposal. This limitation constrains the understanding of whether an artificial intelligence-mediated intervention can simultaneously and coherently favor the physical development and emotional balance of students.

Consequently, the present work aims to evaluate the effect of a physical education program mediated by artificial intelligence on the physical fitness and emotional well-being of basic education students, and to investigate the relationship between physical and emotional gains. The relevance of this proposal lies in providing a replicable methodological model that articulates the objective assessment of physical fitness with the measurement of emotional well-being within the framework of educational technological innovation.

Materials and methods

Study Design

A quasi-experimental design with two non-equivalent groups—experimental and control—and repeated measurements at two time points, before and after the intervention, was adopted, with a duration of twelve weeks. This approach was selected for its feasibility in real school contexts, where individual random assignment is unfeasible and students are grouped into pre-existing classrooms; therefore, assignment to conditions was carried out by clusters (complete classrooms), preserving the natural organization of the class group and reducing contamination between conditions. The unit of analysis was the student, and the factorial design was mixed, combining a between-subjects factor (group: experimental vs. control) and a within-subjects factor (time: pre- vs. post-measurement). To minimize biases, the physical tests and the well-being scale were administered by evaluators blind to the participants' condition (blinded assessment), the volume and frequency of physical education classes were kept constant in both groups, and contextual variables such as session schedules and teacher characteristics were controlled as much as possible. The planning and reporting of the study conformed to recommendations for non-randomized investigations.

Se adoptó un diseño cuasiexperimental con dos grupos no equivalentes —experimental y control— y mediciones repetidas en dos momentos, antes y después de la intervención, con una duración de doce semanas. Esta aproximación se seleccionó por su viabilidad en contextos escolares reales, en los que la asignación

Contexto, población y muestra

The reference population consisted of basic education students from the municipality of San Gil in the Department of Santander, Colombia. The study was conducted in educational institutions in that city over a twelve-week period corresponding to a block of the 2025-2026 school year. The sample included 240 students, evenly distributed between an experimental group (n = 120) and a control group (n = 120), with equal representation from grades 6 and 7 (n = 120) and grades 8 and 9 (n = 120)—60 students per grade level in each group—and a balanced distribution by sex. Group formation sought initial equivalence in educational level, age, and sex composition, a condition that was empirically verified and reported in the results section.

Inclusion criteria: students enrolled in the considered basic education levels, with student assent and informed consent from the responsible adult, and without medical contraindication for the practice of physical activity.

Exclusion criteria: presence of injuries or medical conditions that prevented the performance of physical tests, participation in less than eighty percent of the sessions, or absence from any of the measurements. Sampling is described as non-probabilistic convenience sampling, with group assignment by clusters (classrooms) to reduce contamination between conditions.

Artificial Intelligence-Mediated Intervention

The grupo experimental group developed the physical education program supported by four artificial intelligence-based technological components, while the control group continued with conventional physical education classes of equal frequency and volume. The components were: (1) physical activity monitoring applications with adaptive plans and automated performance feedback, consistent with evidence on wearable devices in the school environment (Casado-Robles et al., 2022; Chen et al., 2025); (2) an educational conversational assistant oriented towards psychoeducation and emotional self-regulation, framed within social and emotional learning programs (Cipriano et al., 2023; Dritsona et al., 2025); (3) automated motor gesture analysis using computer vision for technique correction; and (4) active video games to promote enjoyment and adherence to physical activity (Marsigliante et al., 2024; Rosi et al., 2025). The pedagogical design was grounded in the artificial intelligence application framework in physical education proposed by Lee and Lee (2021) and addressed teaching competencies and artificial intelligence literacy (Ng et al., 2024).

The application of the program was carried out by the physical education teachers, previously trained in the use of the tools and provided with protocols and instructions for each component, to ensure implementation homogeneity. Intervention fidelity was controlled through usage logs of the applications and attendance records, including in the analysis only those students with participation equal to or greater than eighty percent of the sessions. The control group developed conventional physical education classes with the same class load and same curricular content, so that the only systematic difference between conditions was the artificial intelligence-based technological mediation.

Variables and instruments

Cardiorespiratory endurance. The twelve-minute run test (Cooper test) was used, recording the distance covered in meters. From this, maximum oxygen consumption was estimated using the Cooper equation: maximal oxygen consumption = (distance in meters − 504.9) / 44.73. The validity and reliability of this test in preadolescents and adolescents have been the subject of systematic review, so its estimates should be interpreted with caution (Martínez-Lemos et al., 2024); as a well-validated field alternative, the twenty-meter shuttle run test is also recognized (Mayorga-Vega et al., 2015). In its application, the test is conducted on a marked track and consists of covering the greatest possible distance in twelve minutes at a self-regulated pace; the total distance achieved allows estimation of maximal oxygen consumption, a reference indicator of aerobic capacity.

Upper body muscular endurance**.** The elbow push-up test in thirty seconds was applied, counting the number of correct repetitions. The test quantifies upper body muscular endurance based on the maximum number of correct repetitions performed in thirty seconds, according to a standardized technical criterion for movement amplitude and cadence.

Speed. The time, in seconds, taken to cover fifty meters in a straight line was measured, where lower values indicate better performance. Time was recorded using a stopwatch on a fifty-meter straight-line course with a standing start, retaining the best of two attempts; being a speed test, shorter times express better performance.

Emotional well-being. The World Health Organization Well-Being Index, composed of five items and transformed to a scale from 0 to 100, was used, where higher scores reflect better well-being; this is a brief self-report instrument with adequate psychometric properties (Topp et al., 2015). The index comprises five positively connoted statements referring to the last two weeks, rated on an ordinal scale; the raw score is transformed to a range of 0 to 100, where higher values reflect a better state of well-being, and constitutes a brief, sensitive, and easily administered measure.

Procedure

The study was organized into three phases. In the baseline phase (week zero), the four measurements—Cooper test, elbow push-ups, fifty-meter speed, and emotional well-being—were administered under standardized conditions and within the same time slot to all participants in both groups. During the intervention phase (weeks one to twelve), the experimental group developed the artificial intelligence-mediated physical education program, while the control group followed conventional physical education with identical frequency and volume. In the post-intervention phase (week thirteen), the measurements were repeated with the same instruments, the same evaluators, and the same application conditions as in the baseline. To reduce biases, the evaluators were unaware of the condition of each classroom, the order of test application was kept constant, and data were coded anonymously before analysis.

Sample Size and Statistical Power

The sample size (N=240; n=120 per group) was considered adequate to detect effects of moderate magnitude. A sensitivity analysis indicates that, with two groups, a significance level of 0.05, and a power of 0.80, the design allows detecting standardized differences from d $\approx$ 0.37, a value lower than the effects expected according to the literature. The statistical power achieved (post hoc) for each contrast is reported in the results section.

Statistical Analysis

Data management and cleaning were performed using Microsoft Excel, and statistical processing with the R language and environment (version 4.4.1). Data organization and transformation were carried out with the tidyverse ecosystem (dplyr and tidyr packages), and descriptive statistics—means, standard deviations, and intervals—were obtained with the psych package. To ensure reproducibility, a random seed was set (set.seed) before any resampling-based procedure.

The inferential analysis plan comprised: (a) verification of baseline equivalence between groups using independent samples t-tests and the chi-square test for sex; (b) verification of normality assumptions with the Shapiro-Wilk test (shapiro.test function) and homogeneity of variances with Levene's test (car package); (c) an analysis of covariance (ANCOVA) on the post-measurement, with the baseline as covariate, estimated with the car package (type III sums of squares) and whose adjusted marginal means and 95% confidence intervals were obtained with the emmeans package; (d) a mixed two-factor (group × time) analysis of variance with the afex package (aov_ez function) and partial eta squared; (e) t-tests for related and independent samples—the latter with Welch's correction for unequal variances—and, when assumptions were not met, their non-parametric equivalent (Wilcoxon tests); (f) calculation of Cohen's d effect size and partial eta squared with the effect size package, with 95%

confidence intervals obtained by bootstrap resampling (2000 repetitions) using the boot package; (g) statistical power analysis—both a priori sensitivity calculation and achieved power—with the pwr package; (h) moderation analysis, evaluating group × educational level and group × sex interactions on change scores; (i) test-retest stability, estimated with pre-post correlation in the control group; and (j) Pearson correlations between changes in physical fitness and emotional well-being, with their confidence intervals and power, calculated with the correlation package. Figures were generated with the ggplot2 package. A significance level of 0.05 was established. As an essential methodological limitation, it is reiterated that the data are simulated, so the results illustrate the analytical procedure and do not constitute empirical evidence.

Introduction

Los análisis se realizaron sobre la totalidad de los 240 casos simulados, sin valores perdidos. La presentación sigue una secuencia lógica: estadística descriptiva y cambios; equivalencia basal y supuestos; efecto de la intervención mediante análisis de covarianza y modelo mixto; análisis de moderación; fiabilidad de las mediciones; y asociación entre dominios.

The analyses were performed on all 240 simulated cases, with no missing values. The presentation follows a logical sequence: descriptive statistics and changes; baseline equivalence and assumptions; intervention effect via analysis of covariance and mixed model; moderation analysis; reliability of measurements; and association between domains.

***Descriptive Statistics and Pre-Post* Changes**

varianza mixto de dos factores (grupo × tiempo) con el paquete afex (función aov_ez) y eta cuadrado parcial; (e) pruebas t para muestras relacionadas e independientes —estas últimas con corrección de Welch ante varianzas desiguales— y, cuando los supuestos no se cumplieron, su equivalente no paramétrico (pruebas de Wilcoxon); (f) el cálculo del tamaño del efecto d de Cohen y del eta cuadrado parcial con el paquete effectsize, con intervalos de confianza del 95 % obtenidos por remuestreo bootstrap (2000 repeticiones) mediante el paquete boot; (g) el análisis de la potencia estadística —tanto el cálculo a priori de sensibilidad como la potencia alcanzada— con el paquete pwr; (h) el análisis de moderación, evaluando las interacciones grupo × nivel educativo y grupo × sexo sobre las puntuaciones de cambio; (i) la estabilidad test-retest, estimada con la correlación pre-post en el grupo control; y (j) las correlaciones de Pearson entre los cambios de la condición física y del bienestar emocional, con sus intervalos de confianza y potencia, calculadas con el paquete correlation. Las figuras se elaboraron con el paquete ggplot2. Se fijó un nivel de significación de 0,05. Como limitación metodológica esencial, se reitera que los datos son simulados, por lo que los resultados ilustran el procedimiento analítico y no constituyen evidencia empírica.

Introducción

Los análisis se realizaron sobre la totalidad de los 240 casos simulados, sin valores perdidos. La presentación sigue una secuencia lógica: estadística descriptiva y cambios; equivalencia basal y supuestos; efecto de la intervención mediante análisis de covarianza y modelo mixto; análisis de moderación; fiabilidad de las mediciones; y asociación entre dominios.

Estadística descriptiva y cambios pre-post

Table 1.
Descriptive statistics (mean and standard deviation) by group and time point. (N = 240).

Variable	Group	Pre M (SD)	Post M (SD)	Change (Δ)
Cooper Test (m)	Experimental	2013.9 (283.1)	2101.3 (288.5)	+87.4
Cooper (m)	Control	2018.1 (286.5)	2049.3 (291.8)	+31.2
Estimated VO2max (ml/kg/min)	Experimental	33.7 (6.3)	35.7 (6.5)	+2.0
Estimated VO2max (ml/kg/min)	Control	33.8 (6.4)	34.5 (6.5)	+0.7
Elbow Push-ups 30 s (rep)	Experimental	17.1 (4.8)	19.4 (5.3)	+2.4
Elbow Push-ups 30 s (rep)	Control	15.7 (5.9)	16.4 (6.3)	+0.7
Speed 50 m (s)	Experimental	8.88 (1.11)	8.60 (1.15)	−0.28
Speed 50 m (s)	Control	8.77 (1.03)	8.63 (1.05)	−0.14
WHO-5 Well-being (0–100)	Experimental	61.2 (13.3)	67.5 (14.2)	+6.4
WHO-5 Well-being (0–100)	Control	61.4 (13.4)	63.1 (13.8)	+1.7

Examination of the descriptive statistics reveals a consistent pattern of greater improvement in the experimental group. In relative terms, this group increased aerobic capacity by approximately 4.3% (compared to 1.5% for the control), muscular endurance by about 14% (compared to 4.5%), speed by 3.2% (compared to 1.6%), and emotional well-being by 10.4% (compared to 2.8%), while estimated maximal oxygen consumption increased by 5.9% (compared to 2.1%). Intragroup contrasts using related-samples t-tests confirmed that these changes were statistically significant in the experimental group for all four variables (all $p < 0.001$), with large effect sizes (Cohen's d between 1.00 and 1.24); in the control group, changes also reached significance, although with small to moderate magnitudes (Cohen's d between 0.36 and 0.61). This preliminary contrast in the magnitude of changes anticipates the differential effect formally confirmed in subsequent inferential analyses.

After the intervention, the experimental group increased the distance in the Cooper test by 87.4 meters compared to 31.2 for the control, improved by 2.4 repetitions in push-ups compared to 0.7, reduced the fifty-meter speed time by 0.28 seconds compared to 0.14, and increased well-being by 6.4 points compared to 1.7. Figure 1 illustrates the evolution of the means, and Figure 2 shows the complete distributions using box plots, evidencing the favorable shift of the experimental group without influential outliers.

Figure 1. *Means (with standard error) before and after the intervention, by group.*

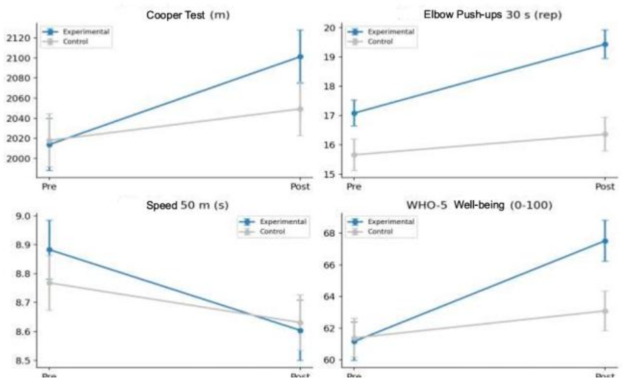
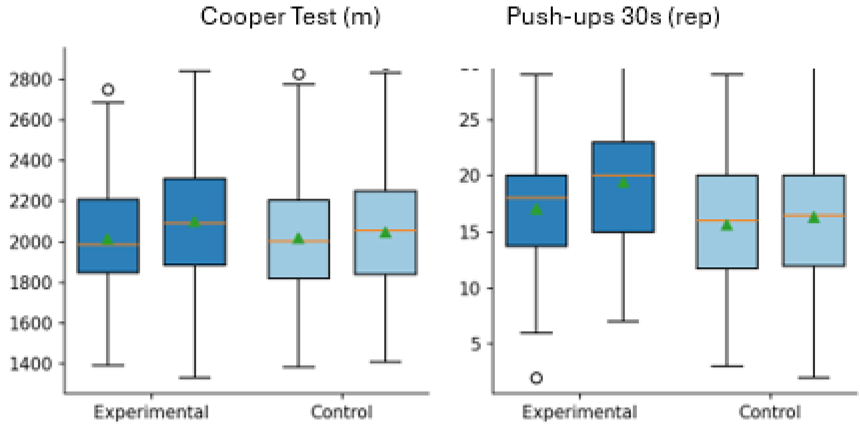


Figure 2. *Distributions (box plots) pre and post by group.*



Baseline Equivalence and Assumption Verification

Before the intervention, the groups did not differ significantly in most variables, with trivial effect sizes (Table 2). Only elbow push-ups showed a small baseline difference (p=0.041; d=0.27), and sex distribution was homogeneous between groups (χ^2=2.40; p=0.121). This equivalence, together with the use of analysis of covariance, controls for the possible selection bias inherent in quasi-experimental designs.

Table 2.

Baseline Equivalence and Assumption Verification

Variable	M Exp.	M Control	t	p	d
Cooper Test (m)	2013.9	2018.1	−0.11	0.910	0.01
Elbow Push-ups 30 s (rep)	17.1	15.7	2.05	0.041	0.27
Speed 50 m (s)	8.88	8.77	0.83	0.409	0.11
WHO-5 Well-being (0–100)	61.2	61.4	−0.13	0.896	0.02

The Shapiro-Wilk test indicated normality of change scores for the Cooper test, speed, and well-being (p>0.05), while push-ups deviated from normality; Levene's test revealed homogeneity of variances for Cooper and well-being, and unequal variances for push-ups and speed (Table 3). Consequently, between-group comparisons were performed with Welch's correction and the main contrasts were complemented with analysis of covariance, robust to these deviations; results were confirmed through non-parametric tests, with equivalent conclusions.

Tabla 3.

Verification of assumptions about change scores

Variable	Shapiro p (Exp.)	Shapiro p (Ctrl.)	Levene p	Decision
Cooper Test (m)	0,206	0,802	0,158	Met
Elbow Push-ups 30 s	0,007	< 0,001	0,002	Welch/non-param.
Speed 50 m (s)	0,296	0,456	0,035	Welch
WHO-5 Well-being	0,409	0,505	0,117	Met

Intervention Effect: ANCOVA and Mixed Model

The analysis of covariance, adjusting the post-measurement for the baseline, confirmed a significant group effect in all four variables, with systematically more favorable adjusted means in the experimental group (Table 4). Convergently, the mixed analysis of variance evidenced significant group × time interactions, and change scores differed between groups with effect sizes ranging from moderate to large, confidence intervals excluding zero, and achieved statistical power close to one (Table 5; Figure 3).

Table 4.
Analysis of covariance (post-measurement adjusted by baseline) and adjusted marginal means with 95% CI.

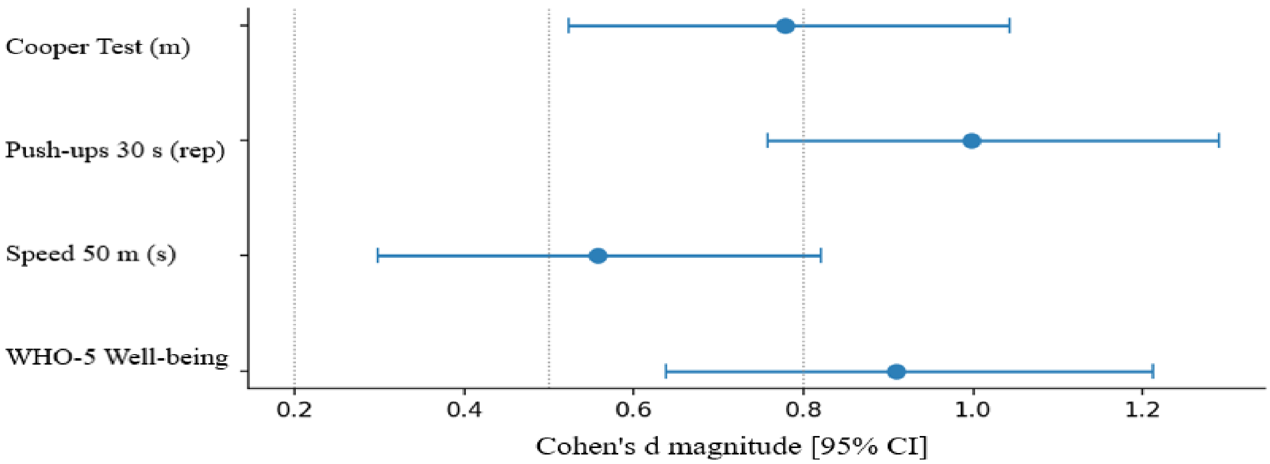
Variable	F(1,237)	η^2 partial	Adj. M Exp. [95% CI]	Adj. M Ctrl. [95% CI]
Cooper Test (m)	36,25	0,133	2103,4 [2090,4–2116,4]	2047,3 [2034,3–2060,3]
Push-ups 30 s (rep)	55,65	0,190	18,70 [18,40–18,99]	17,10 [16,80–17,39]
Speed 50 m (s)	18,38	0,072	8,55 [8,50–8,59]	8,69 [8,64–8,73]
WHO-5 Well-being	49,48	0,173	67,63 [66,70–68,55]	62,98 [62,06–63,90]

All ANCOVA effects were significant ($p<0.001$). Partial eta squared values, between 0.072 and 0.190, correspond to medium to large effects.

Table 5.
Group × time interaction, between-group difference in change, effect size with 95% CI, and power

Variable	Interaction F(1,238) [η^2p]	Cohen's d [95% CI]	Power	p
Cooper Test (m)	36.38 [0,133]	0.78 [0,52–1,04]	1.00	< 0.001
Push-ups 30 s (rep)	59.77 [0,201]	1.00 [0,76–1,29]	1.00	< 0.001
Speed 50 m (s)	18.63 [0,073]	−0.56 [−0,82–−0,30]	0.99	< 0.001
WHO-5 Well-being	49.62 [0,173]	0.91 [0,64–1,21]	1.00	< 0.001

Figure 3. *Between-group effect size (Cohen's d) with 95% confidence intervals.*



Moderation Analysis

Group × educational level and group × sex interactions were not significant in any of the variables (all p > 0.28 and partial eta squared ≤ 0.005), indicating that the magnitude of the intervention effect was homogeneous between primary and secondary levels and between male and female students. This consistency supports the robustness and potential generalizability of the result pattern within the simulated model.

Measurement Reliability

Test-retest stability, estimated by the correlation between pre- and post-measurements in the control group, was high for physical tests (r = 0.98 for the Cooper test, push-ups, and speed) and high for emotional well-being (r = 0.94), suggesting that instruments offered consistent measurements and that observed changes in the experimental group are not attributable to random instrument fluctuations.

Association between Physical Fitness and Emotional Well-being

In the experimental group, changes in physical fitness were associated with improvements in emotional well-being (Table 6; Figure 4). The highest correlation corresponded to the relationship between change in aerobic capacity and change in well-being, followed by strength; improvement in speed (negative change in time) was also associated with greater well-being. The correlation matrix among all changes (Figure 5) shows a coherent pattern of low to moderate magnitude associations.

Table 6.
Pearson correlations between changes in physical fitness and emotional well-being (experimental group, n = 120).

Relationship between changes	r [95 % CI]	p	Power
Δ Aerobic capacity (Cooper) – Δ Well-being	0.38 [0,21–0,52]	< 0.001	0.99
Δ Strength (push-ups)– Δ Well-being	0.28 [0,11–0,44]	0.002	0.88
Δ Speed (50 m) – Δ Well-being	−0.25 [−0,41–−0,08]	0.006	0.80

Figure 4. *Scatter plots with regression line between changes in physical fitness and well-being (experimental group).*

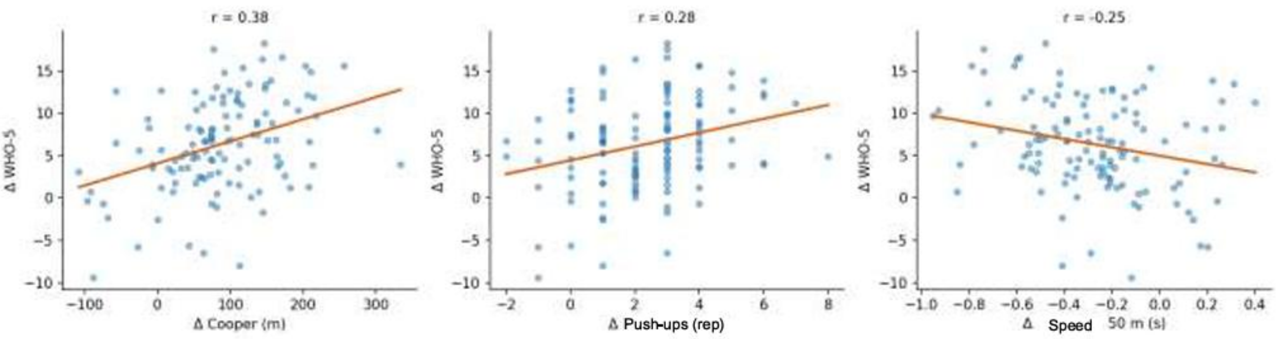
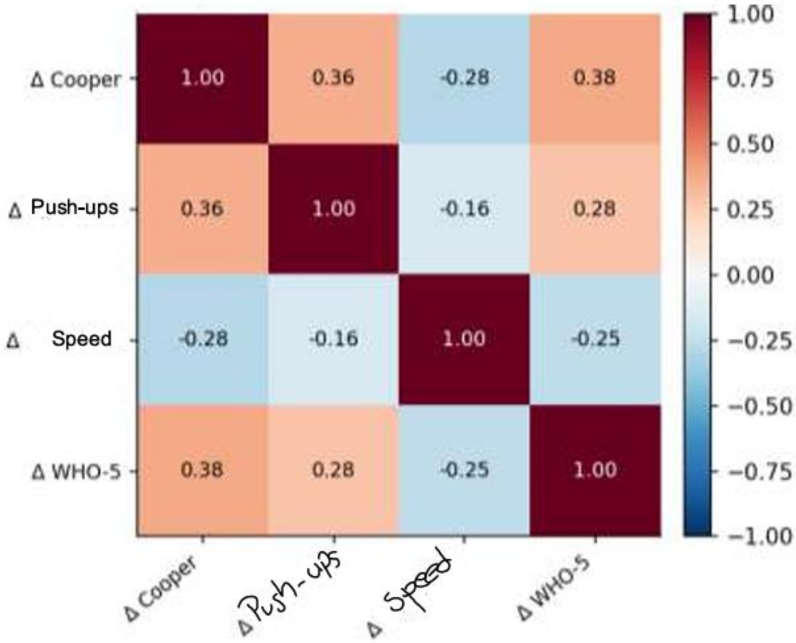


Figure 5. *Correlation matrix of changes (experimental group).*



Discussion

The present study evaluated, using a methodological model with simulated data, the effect of a physical education program mediated by artificial intelligence on the physical fitness and emotional well-being of basic education students. The simulated results showed significantly greater improvements in the experimental group across all four variables, with significant group-by-time interaction effects, favorable adjusted means after controlling for baseline measurement, effect sizes ranging from moderate to large with precise confidence intervals, and positive associations between physical and emotional gains. These findings should be interpreted as an illustration of the expected behavior under the study hypotheses, and not as empirical evidence.

The direction of the simulated results is consistent with the available literature. Regarding physical fitness, the observed favorable effects align with evidence on school-based programs supported by wearable devices, which have reported small to moderate increases in daily physical activity (Casado-Robles et al., 2022), although with heterogeneous results depending on the design and type of device (Chen et al., 2025). Similarly, active video games have shown improvements in physical fitness, enjoyment, and psychological well-being in schoolchildren (Marsigliante et al., 2024; Rosi et al., 2025), supporting the plausibility of the gamification component included in the intervention.

Regarding emotional well-being, the simulated improvement in the experimental group is consistent with the body of evidence linking physical activity with benefits in child and adolescent mental health, particularly on stress and social competence (Fu et al., 2025), and with school-based trials that have documented reductions in depressive symptoms and increases in life satisfaction (Ahmed et al., 2023). The incorporation of a conversational assistant oriented towards emotional self-regulation aligns with the robust meta-analytic evidence on social and emotional learning programs (Cipriano et al., 2023) and with the still emerging and cautious perspectives on the use of artificial intelligence in the mental health of minors (Dritsona et al., 2025).

The positive association between physical and emotional gains in the experimental group suggests, within the limits of the model, that both domains could benefit jointly when artificial intelligence articulates training personalization with socio-emotional support (Lee & Lee, 2021; Li et al., 2024; Zhang et al., 2022). However, an alternative possible explanation is that the improvement in well-being derives from greater enjoyment and technological novelty rather than from the physical gain itself, which should be clarified in empirical studies with mediation analysis.

From the perspective of analytical robustness, the results were consistent across complementary procedures: the analysis of covariance, which controls for baseline measurement, coincided with the mixed model and the comparison of change scores; effects were maintained after applying Welch's correction for unequal variances and non-parametric tests; effect size confidence intervals were precise and achieved statistical power was close to unity; and the moderation analysis did not reveal differences according to educational level or sex, pointing to a stable pattern. The high test-retest stability reinforces measurement reliability. Together, these elements illustrate the type of methodological control required of quality research, although their value remains conditioned by the simulated nature of the data.

Among the strengths of the approach are the integration of two domains usually studied separately, the use of recognized field tests and a validated well-being index, and the application of a comprehensive and reproducible analytical plan. As limitations, the most relevant is that the data are simulated, so conclusions cannot be generalized to real populations; the quasi-experimental design does not allow strong causal inferences; field tests provide estimates and not direct measures, especially the Cooper test in preadolescents, whose validity has been questioned (Martínez-Lemos et al., 2024); and the twelve-week period does not inform about the sustainability of the effects. Likewise, the use of artificial intelligence with minors poses ethical challenges of privacy, equity, and human supervision that condition its implementation (Farrelly & Baker, 2023; Ng et al., 2024).

The implications of this model are primarily methodological and pedagogical. Methodologically, the study provides a replicable template for articulating the objective assessment of physical fitness with the measurement of emotional well-being. Pedagogically, it offers an intervention scheme that combines adaptive monitoring, motor gesture analysis, gamification, and socio-emotional support. Future lines of research should execute the protocol with empirical data, employ cluster-randomized designs with long-term follow-up, incorporate objective measures of physical activity, analyze mediation mechanisms between physical fitness and well-being, and evaluate equity of access to technological tools. In summary, the scope of this work is limited to demonstrating the coherence of an analytical model; its significance will depend on subsequent empirical verification.

Conclusions

In line with the stated objective, the results obtained from simulated data suggest that a physical education program mediated by artificial intelligence could be associated with superior improvements in physical fitness and emotional well-being in basic education students, compared to conventional physical education, and that both improvements would tend to be positively related. These claims are illustrative and cautious in nature, as they are not derived from real measurements.

Therefore, it is concluded that artificial intelligence, used as a pedagogical tool, constitutes a promising pathway to integrate physical development and emotional balance within basic education, provided that its adoption is accompanied by methodological rigor, ethical safeguards, and teacher supervision. Research with empirical data, robust designs, and long-term follow-up is required to confirm these trends and specify the mechanisms involved.

References

Ahmed, K. R., Horwood, S., & Khan, A. (2023). Effects of a school-based physical activity intervention on adolescents' mental health: A cluster randomized controlled trial. Journal of Physical Activity and Health, 20(12), 1102–1108. https://journals.humankinetics.com/view/journals/jpah/20/12/article-p1102.xml

Aubert, S., Barnes, J. D., Demchenko, I., Hawthorne, M., … Tremblay, M. S. (2022). Global Matrix 4.0 physical activity report card grades for children and adolescents: Results and analyses from 57 countries. Journal of Physical Activity and Health, 19(11), 700–728. https://doi.org/10.1123/jpah.2022-0456

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., … Willumsen, J. F. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. British Journal of Sports Medicine, 54(24), 1451–1462. https://doi.org/10.1136/bjsports-2020-102955

Casado-Robles, C., Viciana, J., Guijarro-Romero, S., & Mayorga-Vega, D. (2022). Effects of consumer-wearable activity tracker-based programs on objectively measured daily physical activity and sedentary behavior among school-aged children: A systematic review and meta-analysis. Sports Medicine - Open, 8, 18. https://doi.org/10.1186/s40798-021-00407-6

Chen, X., et al. (2025). Effectiveness of wearable activity trackers on physical activity among adolescents in school-based settings: A systematic review and meta-analysis. BMC Public Health, 25(1). https://doi.org/10.1186/s12889-025-22170-z

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., … Durlak, J. A. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. Child Development, 94(5), 1181–1204. https://doi.org/10.1111/cdev.13968

Dritsona, A., Koutroumpa, M. E., Sergentanis, T. N., & Tsitsika, A. K. (2025). Artificial intelligence and mental health of children and adolescents: Current status and perspectives. Advances in Experimental Medicine and Biology, 1489, 477–484. https://doi.org/10.1007/978-3-032-03394-9_45

Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. Education Sciences, 13(11), 1109. https://doi.org/10.3390/educsci13111109

Fu, Q., Li, L., Li, Q., & Wang, J. (2025). The effects of physical activity on the mental health of typically developing children and adolescents: A systematic review and meta-analysis. BMC Public Health, 25(1), 1514. https://doi.org/10.1186/s12889-025-22690-8

Lee, H. S., & Lee, J. (2021). Applying artificial intelligence in physical education and future perspectives. Sustainability, 13(1), 351. https://doi.org/10.3390/su13010351

Li, S., Wang, C., & Wang, Y. (2024). Fuzzy evaluation model for physical education teaching methods in colleges and universities using artificial intelligence. Scientific Reports, 14. https://doi.org/10.1038/s41598-024-53177-y

Marques, A., Henriques-Neto, D., Peralta, M., Martins, J., Demetriou, Y., Schönbach, D. M. I., & Gaspar de Matos, M. (2020). Prevalence of physical activity among adolescents from 105 low, middle, and high-income countries. International Journal of Environmental Research and Public Health, 17(9), 3145. https://doi.org/10.3390/ijerph17093145

Marsigliante, S., My, G., Mazzotta, G., & Muscella, A. (2024). The effects of exergames on physical fitness, body composition and enjoyment in children: A six-month intervention study. Children, 11(10), 1172. https://doi.org/10.3390/children11101172

Martínez-Lemos, I., et al. (2024). Reliability and criterion-related validity of the Cooper test in pre-adolescents and adolescents: A systematic review and meta-analysis. Journal of Sports Sciences, 42(3), 222–236. https://doi.org/10.1080/02640414.2024.2326352

Mayorga-Vega, D., Aguilar-Soto, P., & Viciana, J. (2015). Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. Journal of Sports Science & Medicine, 14(3), 536–547. https://www.jssm.org/hfpdf.php?volume=14&issue=3&page=536

Ng, D. T. K., Su, J., Leung, J. K. L., & Chu, S. K. W. (2024). Artificial intelligence (AI) literacy education in secondary schools: A review. Interactive Learning Environments, 32(10), 6204–6224. https://doi.org/10.1080/10494820.2023.2255228

Rosi, E., Bianchi, V., Baù, I., Nuzzo, R., Valsecchi, S., Molteni, M., & Colombo, P. (2025). Effects of exergames on motor skills, psychological well-being, and cognitive abilities in schoolchildren and adolescents: Scoping review. JMIR Pediatrics and Parenting, 8, e71416. https://doi.org/10.2196/71416

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. Psychotherapy and Psychosomatics, 84(3), 167–176. https://doi.org/10.1159/000376585

van Sluijs, E. M. F., Ekelund, U., Crochemore-Silva, I., Guthold, R., Ha, A., Lubans, D., et al. (2021). Physical activity behaviours in adolescence: Current evidence and opportunities for intervention. The Lancet, 398(10298), 429–442. https://doi.org/10.1016/S0140-6736(21)01259-9

Xu, K., et al. (2023). Predicting academic performance associated with physical fitness of primary school students using machine learning methods. Complementary Therapies in Clinical Practice, 51, 101736. https://doi.org/10.1016/j.ctcp.2023.101736

Zhang, B., Jin, H., & Duan, X. (2022). Physical education movement and comprehensive health quality intervention under the background of artificial intelligence. Frontiers in Public Health, 10, 947731. https://doi.org/10.3389/fpubh.2022.947731

Declaration of conflicts of interest:

Through this document, the authors of this article assume the regulatory and normative provisions of the RIAF Journal's publication and therefore declare no conflicts of interest in this regard.

Authors' Contribution:

Jorge Andrés Bayona Martínez: Elaborated the research methodology, developed the diagnostic instruments and techniques, and carried out the application of the program in the experimental group.

Edwin Esneider Andrade Perdomo: Conducted the bibliographic search and construction of citations, references, and theoretical systematization of the study variables.

Steven Arturo Torres Burgos: Participated in the review, analysis, and decision-making in the writing of the literature, theoretical systematization of the study variables, application of empirical instruments, and in the results and conclusions of the research.