

RESEARCH ARTICLE

Integrating physiological data monitoring into physical instructional optimization on fatigue state and physical fitness

Jingbo Wang*

Department of Physical Education, Anhui Wenda University of Information Engineering, Hefei, Anhui, China.

Received: April 24, 2026; accepted: August 8, 2026.

Adolescent physical fitness decline and cumulative fatigue pose persistent challenges in physical education. Despite the proliferation of wearable devices, a disconnect between data collection and instructional decision-making creates a "data-information-decision" island, hindering optimized training. This study applied a 12-week, 2×2 factorial randomized controlled trial (RCT) to evaluate the effectiveness of integrating physiological monitoring into physical education among 584 secondary school students aged 13 – 17. Participants engaged in three 40-minute sessions weekly and were allocated into four groups of control, passive monitoring, teacher feedback, and dual feedback. Multimodal data including heart-rate variability (HRV) indexed by the root mean square of successive differences (RMSSD) and the standard deviation of normal-to-normal intervals (SDNN), rating of perceived exertion (RPE), salivary cortisol, maximal oxygen uptake, and shuttle-run scores were collected. The results indicated significant group differences ($P < 0.01$), exhibiting a clear dose-response gradient. The dual feedback group achieved the largest improvements with a 32.4% reduction in the fatigue composite index and a 24.7% increase in the fitness composite index, while the teacher-feedback group improved by 21.8% and 17.3%, the passive-monitoring group by 8.6% and 6.2%, and the control group by 3.2% and 2.8%, respectively. Mediation analysis revealed two pathways including "monitoring data → instructional personalization and load adequacy → fatigue relief" with 48.3% mediation and "monitoring data → feedback timeliness and training precision → fitness gain" with 51.7% mediation. Heterogeneity analysis indicated that students with lower baseline fitness and higher device adherence benefited most. Integrating physiological monitoring into pedagogical decision-making was essential for optimizing outcomes. The proposed "Diagnosis–Design–Implementation–Evaluation" framework provided a roadmap for transforming physical education from intuition-based to precision-based practice.

Keywords: physiological monitoring; fatigue; physical fitness; randomized controlled trial; instructional optimization; data-driven decision-making.

*Corresponding author: Jingbo Wang, Department of Physical Education, Anhui Wenda University of Information Engineering, Hefei, Anhui 231201, China. Email: 3308342@163.com.

Introduction

The landscape of school physical education is being reshaped by the dual global momentum of deepening health-education initiatives and prioritizing adolescent well-being [1]. Classrooms are shifting from traditional activity-based

models toward data-informed interaction, integrating wearable sensors, real-time biofeedback, and digital analytics to enrich training experiences [2]. This transition reflects a paradigm shift in which physical education is valued not only for motor-skill acquisition but also as a core component of holistic student

development, addressing declining cardiorespiratory fitness, rising obesity, and cumulative fatigue among secondary-school populations [1].

Significant theoretical and empirical advances have clarified the physiological basis of exercise monitoring. Wearable sensor technologies now generate continuous, objective profiles of training load and recovery, principally through heart-rate variability, electrodermal activity, and movement acceleration [3]. Concurrent research on data-driven decision-making demonstrates that algorithmic feedback loops can materially improve the precision of coaching interventions, moving practice beyond intuitive adjustments [4]. Studies further confirm that structured physical activity enhances psychological resilience and alleviates performance-related anxiety by strengthening autonomic regulatory capacity and stress-recovery cycles [5]. These developments provide a robust evidence base for linking technological integration to improved physiological outcomes in school settings [6]. Despite these advances, a structural contradiction persists. Wearable devices proliferate, yet the translation of raw physiological data into meaningful pedagogical action remains nascent [7]. High-frequency data streams are routinely captured but confined to devices or cloud platforms, producing entrenched data silos and decision stagnation [8]. Teachers, confronted with unprocessed metrics, face increased cognitive load and decision fatigue, rather than clearer instructional judgement [7], while students, lacking intelligible feedback channels, cannot convert physiological signals into self-regulation [9], which results in a widening gap between technological capacity and educational benefit. Critically, existing work has neither finely quantified how different instructional modalities, varying in data access and feedback use, shape dual-dimension outcomes of fatigue and fitness, nor how these effects vary across student subgroups that demands urgent clarification.

This study proposed a data collection–pedagogical mediation–performance output three-stage chain of effects to systematically examine the differential pathways of four instructional modalities including control, passive monitoring, teacher feedback, dual feedback on composite fatigue and fitness indices. Stratified heterogeneity testing was implemented to reveal how effects vary across subgroups defined by baseline fitness and device adherence, thereby proposing a classroom-optimization roadmap grounded in precise regulation. This research provided a quantifiable framework aligning exercise-science theory with educational practice. By decomposing the transmission pathways between physiological monitoring and dual-dimension outcomes, the study moved beyond generic pedagogical advice to evidence-based strategies for curriculum design. The findings offered the scientific community a replicable paradigm for integrating data-driven decision-making into physical education, supporting training protocols that fostered athletic competence while buffering students against physiological distress. The proposed diagnosis design implementation evaluation closed loop framework supplied a scalable route for transitioning school physical education from intuition-based to precision-based practice.

Materials and methods

Research framework and mechanism

This study constructed a three-stage chain of effects linking physiological data acquisition, pedagogical decision adjustment, and dual-dimension outcome response. Physiological data acquisition was conducted by using Polar H10 chest-strap heart-rate monitor (Polar Electro Oy, Kempele, Finland) to capture high-fidelity electrocardiographic signals for heart-rate variability (HRV) analysis including the root mean square of successive differences (RMSSD) and the standard deviation of normal-to-normal intervals (SDNN). Garmin Forerunner 265 wrist-based optical heart-rate sports watch (Garmin Ltd., Olathe, Kansas, USA) was used to estimate

training load and recovery metrics *via* photoplethysmography. Empatica E4 multimodal wrist-worn physiological sensor (Empatica Srl, Milano, Italy) was used to integrate electrodermal activity (EDA), skin temperature, and tri-axial acceleration to infer autonomic stress responses. Salivary cortisol sampled before and 45 minutes after each session was measured using Salimetrics SCL-600 Salivary Cortisol Enzyme Immunoassay Kit (Salimetrics LLC, State College, Pennsylvania, USA) to supplement wearable streams by indexing endocrine stress-recovery dynamics at the cortisol level. Pedagogical decision adjustment was performed by translating raw physiological streams into actionable instructional decisions. This stage comprised four core elements including intensity regulation, pace restructuring, tiered guidance, and fatigue alerting. Decision adjustment was conducted through feedback loops that differed across the four study groups. No data entered the instructional flow in the control group, while data were recorded but not used for adjustment in the passive monitoring group. In the teacher feedback group, data were pushed to the instructor's interface, prompting real-time modulation of intensity and pacing, while data were concurrently available to both teacher and student in the dual feedback group, enabling collaborative load regulation. The quality and timeliness of this stage determined whether technological capacity was converted into educational benefit rather than remaining trapped in data silos. In dual-dimension outcome response stage, the downstream effects were manifested as relief of fatigue indexed by subjective rating of perceived exertion (RPE), HRV recovery, cortisol dynamics, reaction time and enhancement of physical fitness indexed by maximal oxygen uptake, shuttle-run performance, muscular endurance, flexibility, explosive power. The three stages formed a progressive transmission chain as external physiological signals → central decision-making → bidirectional physiological outcomes. The decision-adjustment stage functioned as the pivotal hub, while its configuration across groups generated the observed heterogeneity in

intervention effects. The four instructional modalities therefore differed systematically along two dimensions of data access and feedback use, producing graded pathways through the same mechanistic chain. This structure supplied the basis for estimating direct and indirect effects *via* empirical models.

Participants and group design

A total of 584 participants with 304 males and 280 females were recruited from four secondary schools in Nanjing, Jiangsu, China including Nanjing No. 29 Junior High School (150 students with 78 males and 72 females), Nanjing No. 1 Junior High School (135 students with 69 males and 66 females), Nanjing Zhonghua Senior High School (160 students with 82 males and 78 females), and Nanjing Normal University Affiliated Senior High School (139 students with 75 males and 64 females). All participants were full-time students and aged between 13 and 17 years old. The inclusion criteria specified no major sports injury within the three months preceding this study, no diagnosed cardiovascular or metabolic disease, no long-term medication affecting heart rate or fatigue assessment, and full capacity to participate in all scheduled sessions and assessments. The exclusion criteria comprised acute or chronic illness episodes, constitutional abnormalities, anticipated attendance instability due to family circumstances or other reasons, and allergic reactions to electrodermal sensors or chest-strap materials. A 2 × 2 factorial randomized controlled trial (RACT) was conducted, crossing two independent factors of monitoring access (access vs. no access) and feedback use (use vs. no use), which yielded four instructional groups with each comprising 146 students as conventional control group, passive-monitoring group, teacher-side feedback group, and dual teacher–student feedback group. Following random assignment, groups were pair-balanced on key dimensions including grade composition, gender ratio, baseline RPE, baseline maximal oxygen uptake ($\text{VO}_{2\text{max}}$), and baseline heart-rate variability (HRV-RMSSD), thereby ensuring initial comparability.

Multimodal physiological data were collected including heart-rate variability and salivary cortisol indexed autonomic stress and recovery [10], maximal oxygen uptake and shuttle-run performance captured cardiorespiratory fitness [10, 12]. Wearable sensors and standardized fitness tests provided objective measurement, data acquisition adhered to a unified protocol under controlled environmental conditions. Structural equation modelling was employed to estimate direct and indirect effects with mediation pathways tested *via* Bootstrap resampling (5,000 replicates) [11]. The intervention lasted 12 weeks with three 40-minute physical education sessions per week including aerobic conditioning, skill drills, and game play. Outcome measures were assessed at five time points including baseline, mid-test 1 (week 4), mid-test 2 (week 8), post-test (week 12), and delayed post-test (week 16). The final point served to evaluate the persistence and decay of intervention effects. All research procedures were approved by the Institutional Review Board (IRB) of Anhui Wenda University of Information Engineering, Hefei, Anhui, China (Approval No. 2025-ETH-011). Written informed consents were obtained from all participants and their legal guardians.

Intervention protocol and implementation

The four study protocols were developed into standardized lesson plans through expert-panel review and pilot-phase calibration, ensuring curricular fidelity across groups. Participants were randomly assigned at the class level to one of four instructional modalities. Control group (G1) had no wearable devices. Physical education proceeded according to the standard school curriculum with teachers adjusting intensity and pacing based on conventional observation. Exercise intensity was targeted at 55 – 70% of maximal heart rate, representing prevailing practice to represent baseline instructional conditions. Passive-monitoring group (G2) wore chest straps and wristbands during all sessions. Physiological data were continuously streamed to a cloud-based management system, stored but not made visible to teachers or students.

Teachers continued to deliver lessons according to the standard curriculum, unaware of real-time data, which isolated the Hawthorne (placebo) effect of device wearing itself, decoupled from any data-driven instructional adjustment. Teacher-feedback group (G3) also wore the same wearable devices to test instructor-centric data use. Physiological data were displayed in real time to the teacher *via* a dashboard interface. Teachers used these data to adjust lesson intensity, pacing, and recovery periods, targeting the 55 – 70% heart-rate zone. Students received no direct feedback. In dual-feedback group (G4), wearable data were streamed to both teacher and student interfaces. Teachers adjusted instructions, while students observed their own heart-rate zones and cumulative load, modulating personal effort accordingly. This experiment examined the combined effect of data-informed teaching and student self-regulation. The lesson duration and curricular content were held constant with only differences in data visibility and use. All groups were taught by the same team of instructors with dual backgrounds in physical education and exercise science, who received three weeks of unified training before the experiment with the first two weeks covering the operational requirements of the standardized lesson plans and equipment-use protocols and the third week involving simulated classes and data-dashboard operation drills. Class recordings were sampled weekly by a third-party expert panel that evaluated four dimensions of procedural fidelity of instructional actions, exercise-load attainment rate, timeliness of decision responses, and student engagement to ensure high intervention fidelity. Student devices recorded data in real time during each session and uploaded post-class to a unified management platform, allowing the instructional team to perform fine-grained adjustments for the following week's sessions according to actual load.

Variable measurement and data collection

Subjective fatigue was captured with the Borg CR10 RPE (<https://www.borgperception.se/>) and completed immediately after each session with

higher scores indicating greater perceived exertion. Objective fatigue was indexed by HRV, specifically RMSSD and SDNN, derived from 10-minute continuous heart-rate recordings under fasted, early-morning resting conditions. Salivary cortisol was assessed with samples being collected before class and 45 minutes after class, and the pre-to-post ratio was computed as an index of stress recovery. Reaction time was measured *via* the Eriksen flanker task implemented using E-Prime 3.0 (<https://pstnet.com/products/e-prime/>) pre- and post-session. The measurement of fitness development spanned four dimensions including cardiorespiratory capacity, muscular strength, flexibility, and speed. Physical fitness was indexed by VO_2max estimated *via* the 20-metre progressive shuttle-run protocol, core muscular endurance (60-second sit-up test), flexibility (sit-and-reach test), and lower-limb explosive power (standing long jump). All fitness assessments followed the National Student Physical Fitness Standards (Ministry of Education of the People's Republic of China, Beijing, China). Control variables included sex, age, family exercise atmosphere, daily screen time, daily sleep duration, dietary structure, and menstrual cycle, acquired through questionnaires and corroborated with parents and homeroom teachers. All indicators were collected at the five assessment points by a unified team following a standardized protocol with ambient temperature held between 22 and 25°C, sessions fixed between 09:00 and 11:00, vigorous exercise prohibited in 24 hours prior to testing, and fasting required for four hours before testing.

Statistical analysis

SPSS 27.0 (IBM, Armonk, NY, USA), R 4.3.0 (<https://www.r-project.org/>), and Mplus 8.9 (<https://www.statmodel.com/>) were employed for the statistical analysis of this study. ANOVA was used to test the main effects of time and group and the time $\times$ group interaction, while time was defined as a within-subjects factor at five points, and group was a between-subjects factor in four groups. Standard degrees-of-freedom tests were applied when the

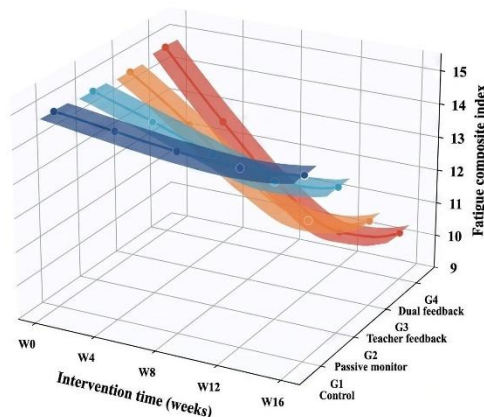
sphericity assumption held, otherwise the Greenhouse–Geisser correction was used. Analysis of covariance (ANCOVA) was additionally employed with baseline scores being entered as covariates to estimate adjusted between-group mean differences on post-test indicators and control for initial baseline variance. Multilevel linear model (MLM) accounted for the nested structure of students within classes with random intercepts at the class level to control for inflation of standard errors due to within-class clustering. Structural equation modelling (SEM) was used to elucidate mediation pathways, through which physiological monitoring affected dual-dimension outcomes with two channels being specified as monitoring intensity $\rightarrow$ instructional personalization, exercise-load adequacy $\rightarrow$ fatigue relief and monitoring intensity $\rightarrow$ feedback timeliness, training precision $\rightarrow$ fitness gain. Mediation significance was judged by 95% confidence intervals derived from 5,000 Bootstrap resamples. Indirect effects with confidence intervals excluding zero were considered statistically significant. Effect sizes were reported using Cohen's d and partial η^2 . Thresholds for Cohen's d were 0.2 (small), 0.5 (medium), and 0.8 (large). The thresholds for partial η^2 were 0.01 (small), 0.06 (medium), and 0.14 (large). Pairwise between-group comparisons applied Bonferroni correction to control Type I error inflation under multiple testing. Heterogeneity analyses were conducted along three dimensions of gender, baseline fitness level, and monitoring compliance, estimating group effects within each subgroup and testing between-subgroup differences. P value less than 0.05 was considered statistically significant difference.

Results and discussion

Descriptive statistics and baseline balance

563 of 584 participants (96.4%) completed all five assessment time points. Baseline characteristics were rigorously balanced across four groups with no significant differences. Multiple imputation

(a) Longitudinal fatigue trajectories across four groups



(b) Distributional gains in fitness composite by groups

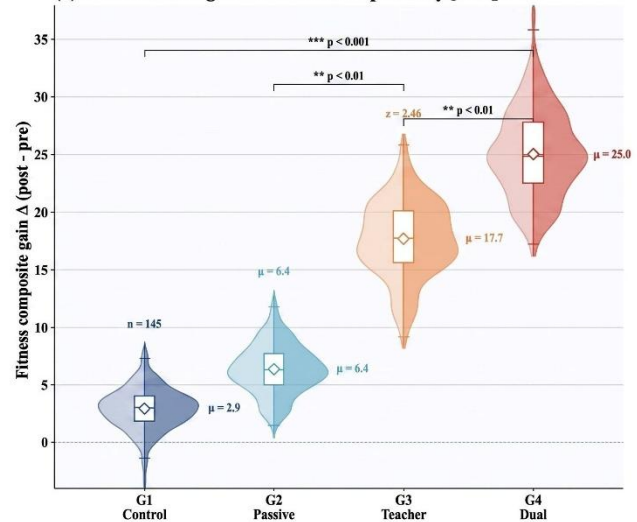


Figure 1. Temporal evolution of fatigue and fitness across the four groups.

was applied for missing data. At baseline, the sample mean subjective RPE was 14.74 ± 1.63 , HRV-RMSSD was 38.47 ± 8.92 ms, $VO_2\text{max}$ was 42.18 ± 5.47 mL/kg/min, and the 20-metre shuttle-run score was 52.36 ± 11.28 levels. Distributions approximated normality, except for the cortisol pre-to-post ratio, which exhibited mild right skew and was Box-Cox transformed prior to analysis.

Overall evaluation of intervention effects

The results indicated that all key indicators exhibited significant main effects of time and group together with a highly significant time $\times$ group interaction. For the composite physiological index, the interaction showed $F = 16.42$ with $P < 0.001$ and partial $\eta^2 = 0.312$, reaching a large-effect magnitude and indicating that the four study groups produced clearly differentiated trajectories during the intervention. With respect to the main effect of time, all groups showed some degree of improvement, but the magnitudes diverged in a marked stepwise fashion. G1 group experienced a 3.2% decline in the fatigue composite and a 2.8% gain in the fitness composite, essentially within the range of natural fluctuation. G2 group declined 8.6% and gained 6.2%, reflecting a mild Hawthorne effect from device wear. G3 group

showed declines of 21.8% and gains of 17.3%, demonstrating a substantial jump once data entered the decision flow G4 group reached 32.4% and 24.7%, respectively, securing further gains on top of the teacher-feedback baseline and corroborating the unique value of the bidirectional feedback mechanism. The longitudinal trajectories of fatigue and fitness composites across the four groups demonstrated a significant time $\times$ group interaction for both fatigue ($F = 16.42$, $P < 0.001$, partial $\eta^2 = 0.312$) and fitness ($F = 14.87$, $P < 0.001$, partial $\eta^2 = 0.289$). Pairwise contrasts (Bonferroni-adjusted) showed that the dual feedback group differed significantly from all other groups on both outcomes at week 12 ($P < 0.001$). The teacher feedback group differed significantly from control and passive monitoring ($P < 0.01$). Passive monitoring differed marginally from control group with $P = 0.042$ for fatigue and $P = 0.051$ for fitness (Figure 1). The results confirmed a dose-response gradient and aligned with increasing data-use intensity.

Intervention effects on fatigue indicators

Fine-grained decomposition of fatigue indicators revealed consistent between-group differentiation. Subjective exertion RPE fell most sharply in the dual-feedback group from 14.78 to

Table 1. Fatigue indicator changes by group (mean $\pm$ SD, % change).

Group	RPE		Δ RPE (Cohen's d)	HRV-RMSSD (ms)		Δ HRV (ms)	Cortisol (%)	Reaction time (ms)
	Baseline	Week 12		Baseline	Week 12			
Control	14.72 $\pm$ 1.61	14.27 $\pm$ 1.58	-0.45 (0.09)	38.51 $\pm$ 8.90	39.66 $\pm$ 8.75	+1.15	3.2	8
Passive Monitoring	14.68 $\pm$ 1.63	13.42 $\pm$ 1.59	-1.26 (0.31)	38.44 $\pm$ 8.94	42.11 $\pm$ 8.82	+3.67	8.6	18
Teacher Feedback	14.81 $\pm$ 1.65	11.58 $\pm$ 1.60	-3.23 (0.62)	38.52 $\pm$ 8.91	48.34 $\pm$ 8.78	+9.82	21.8	32
Dual Feedback	14.78 $\pm$ 1.62	10.18 $\pm$ 1.57	-4.60 (0.95)	38.47 $\pm$ 8.92	52.67 $\pm$ 8.70	+14.14	32.4	48

Notes: All between group differences at week 12 were significant at $P < 0.001$ except control vs. passive monitoring ($P < 0.05$).

Table 2. Fitness outcomes and mediation proportions.

Group	Baseline VO ₂ max	Week 12 VO ₂ max	Δ VO ₂ max (%)	Shuttle-run Δ (levels)	Sit-up Δ (reps)	Fitness gain (%)
Control (G1)	42.21 $\pm$ 5.45	43.03 $\pm$ 5.41	+2.8	+0.54	+1.2	+2.8
Passive monitoring (G2)	42.15 $\pm$ 5.48	44.39 $\pm$ 5.37	+6.2	+1.32	+2.8	+6.2
Teacher feedback (G3)	42.24 $\pm$ 5.46	49.47 $\pm$ 5.32	+17.3	+3.67	+5.1	+17.3
Dual feedback (G4)	42.18 $\pm$ 5.47	52.93 $\pm$ 5.28	+24.7	+5.14	+7.87	+24.7

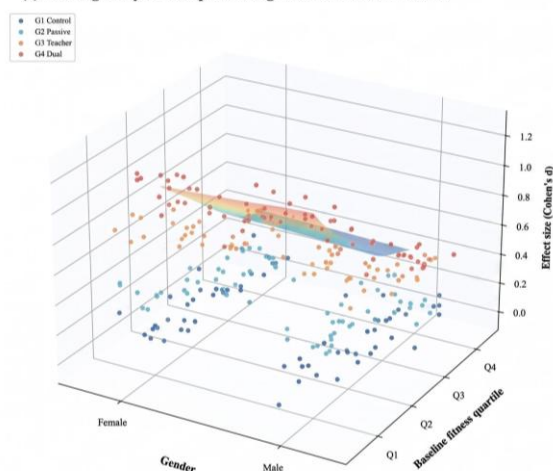
Notes: Mediation: instructional personalization / load adequacy = 48.3%. Feedback timeliness / training precision = 51.7%. All group differences at week 12 were $P < 0.001$.

10.18 ($\Delta = -4.60$, Cohen's $d = 0.95$), approaching a large effect. The teacher-feedback group declined to 11.58 ($\Delta = -3.23$, $d = 0.62$), the passive-monitoring group to 13.42 ($\Delta = -1.26$, $d = 0.31$), and the control group to 14.27 ($\Delta = -0.45$, $d = 0.09$). HRV-RMSSD, indexing vagal-mediated recovery, rose most in the dual-feedback group (+14.14 ms) followed by teacher-feedback (+9.82 ms), passive-monitoring (+3.67 ms), and control (+1.15 ms). The results indicated progressively stronger restoration of autonomic regulatory capacity as data entered instructional decisions. Salivary cortisol pre-to-post ratios declined 20.5% in the dual-feedback group, evidencing accelerated stress recovery. Corresponding reductions were smaller in teacher-feedback (21.8%), passive-monitoring (8.6%), and control (3.2%), respectively. Reaction time shortened most in the dual-feedback group (-48 ms), reflecting faster central-nervous-system recovery post-exertion (Table 1). Collectively, these patterns confirmed that device wearing alone yielded limited benefit, whereas active incorporation of physiological data into pedagogical choice drove meaningful fatigue relief. The dual-feedback configuration, by engaging both teacher and student, produced the most consistent attenuation of physiological strain.

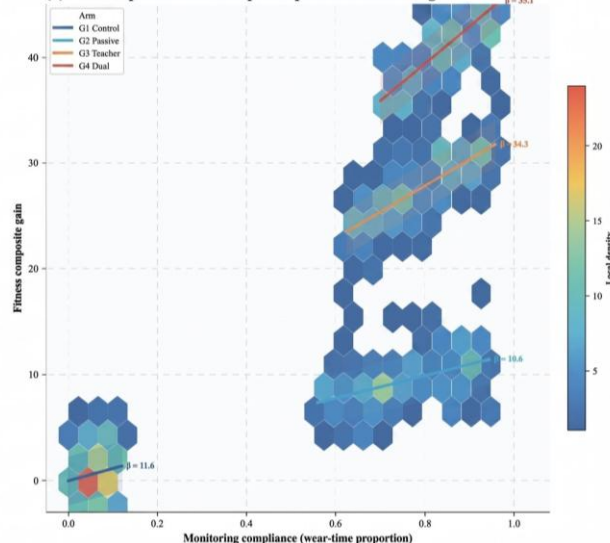
Fitness outcomes and mediation mechanism

Fitness gains paralleled the fatigue trajectories, reinforcing the synergistic action of data-driven regulation across dual outcomes. Maximal oxygen uptake increased by 4.68 units (24.7%) in the dual-feedback group (Cohen's $d = 0.89$) compared with 3.15 units (17.3%) in teacher-feedback, 1.24 units (6.2%) in passive-monitoring, and 0.58 units (2.8%) in control group. The 20-metre shuttle-run performance followed the same gradient with dual-feedback gaining 5.14 levels, teacher-feedback 3.67, passive-monitoring 1.32, and control 0.54, respectively. Core muscular endurance (60-second sit-up test) and associated fitness dimensions showed analogous stepwise differentiation (Table 2). Mediation analysis clarified the pathways translating monitoring into outcomes including monitoring intensity $\rightarrow$ instructional personalization and load adequacy $\rightarrow$ fatigue relief (48.3% of total effect) and monitoring intensity $\rightarrow$ feedback timeliness and training precision $\rightarrow$ fitness gain (51.7% of total effect). Bootstrap 95% confidence intervals (5,000 resamples) excluded zero for all pathway coefficients, confirming statistical significance. The decision-adjustment stage thus functioned as the pivotal hub converting technological capacity into physiological benefit. Without

(a) Heterogeneity landscape across gender and baseline strata



(b) Dose-response relationship: compliance and fitness gain

**Figure 2.** Heterogeneity pattern and compliance–response dose effect.

data-informed adjustment, monitoring remained inert with bidirectional feedback to realize its regulatory potential.

Heterogeneity analysis

Heterogeneity analyses conducted along three dimensions of gender, baseline fitness, and monitoring compliance systematically revealed response differences across subgroups. With respect to gender, males showed slightly larger gains in fitness indicators. In the dual-feedback group (G4), males increased maximal oxygen uptake by 5.12 units compared to 4.23 units among females. By contrast, females exhibited more pronounced improvements in fatigue indicators with G4 group gaining 15.32 ms in HRV-RMSSD, while males gained 13.18 ms. This pattern suggested that elements of playfulness and collaborative participation, inherent to the dual-feedback configuration, more effectively elicited engagement among females, while males derived comparatively greater fitness adaptation. These gender-linked differences were consistent with established variations in exertion perception and motivational orientation during adolescence. The low-baseline subgroup demonstrated the largest relative improvements with the relative change rates reaching 1.4 times

of those in the high-baseline subgroup, indicating that the marginal effect of data-driven instruction was most pronounced where initial capacity was lower. This finding aligned with classical exercise-adaptation theory, wherein physiological returns diminished as baseline fitness raised. Monitoring compliance further amplified these effects. The high-compliance subgroup defined by wear-time proportions above 85% exceeded the low-compliance subgroup with wear-time < 60% by roughly 62% in fitness-composite gain. A clear dose-response relationship thus emerged as greater adherence to device wearing translating into stronger intervention effects, particularly within the dual-feedback condition where data visibility was highest. A three-dimensional scatter and response-surface representation of intervention effects across baseline fitness and gender showed that a pronounced bimodal structure appeared in Cohen's *d* with gains among both males and females in the lowest fitness quartile (Q1) exceeding 1.0, approaching the large-effect threshold. Effect size declined monotonically with rising baseline fitness, falling to approximately 0.45 at the highest quartile (Q4). This gradient confirmed that intervention resources yielded the greatest marginal return

for students with lower initial fitness. A hexagonal density scatter of monitoring compliance against fitness-composite gain showed a positive dose-response relation. Regression slopes increased progressively from 11.6 in the control group (G1) to 35.1 in the dual-feedback group (G4), indicating that compliance amplified intervention effects most strongly where feedback intensity was the highest (Figure 2). The results demonstrated that both individual capacity and behavioral adherence shaped the translational value of physiological monitoring.

Conclusion

This study proposed a diagnosis - design - implementation - evaluation closed - loop framework for data-driven classroom regulation. The diagnosis phase profiled students according to baseline physiological and demographic characteristics, stratifying them by fitness level. The design phase assigned differentiated monitoring-feedback configurations to these strata with the high-fitness tier emphasizing dual feedback, the medium-fitness tier being anchored in teacher feedback, and the low-fitness tier transitioning gradually from passive monitoring to avoid information overload. The implementation phase constructed a periodized training plan that articulated weekly load gradients and feedback density, calibrated to stratum-specific response tendencies. The evaluation phase drew on multimodal physiological data to track intervention effects in near real time, instituting a weekly iterative optimization mechanism. This framework offered a practicable pathway for shifting school physical education from intuition-based to outcome-driven, precision-based practice, aligning exercise-science theory with everyday classroom instruction. The research demonstrated that integrating physiological monitoring into school physical education reduced student fatigue and enhanced physical fitness, but only when monitoring data were actively embedded in pedagogical

decision-making. Device wearing alone yielded limited Hawthorne effects. Bidirectional teacher-student feedback produced the largest gains, outperforming teacher-only feedback and passive monitoring. Two mediating pathways were identified as instructional personalization and load adequacy and feedback timeliness with training precision. The configuration of data visibility and use thus operated as the pivotal hub, translating technological capacity into physiological benefits that concentrated among students with lower baseline fitness and higher device adherence, revealing a clear dose-response relation between monitoring engagement and outcome magnitude. The findings supported a diagnosis - design - implementation - evaluation closed - loop framework for data-driven physical education, which shifted practice from intuition-based instruction toward precision-based regulation, aligning exercise-science theory with classroom delivery. Its scalability rested on three conditions including routine use of validated wearable sensors, transparent feedback channels for both teachers and students, and periodic reprofiling of student physiological states to sustain adaptive adjustment. Limitations of this study included single-city sampling, a 12-week intervention window, and reliance on wearable-derived physiological indicators without deeper biochemical markers. Generalizability to diverse regional and socioeconomic school contexts required further verification. Future work should extend follow-up duration, incorporate metabolic and inflammatory biomarkers, and test implementation across heterogeneous school systems.

Acknowledgements

This work was supported by Key Project of Humanities and Social Sciences in Anhui Province in 2023 (Grant No. 2023AH052810).

References

1. Zhong Q, Jiang J, Bai W, Yin Z, Liao Z, Zhong X. 2025. Application of digital-intelligent technologies in physical education: A systematic review. *Front Public Health*. 13:1626603.
2. Chen H, Xu J, Xie H, Huang Y, Shen X, Xu F. 2022. Effects of physical activity on heart rate variability in children and adolescents: A systematic review and meta-analysis. *Pediatr Exerc Sci*. 34(5):1827–1842.
3. Welch V, Wy TJ, Ligezka A, Hassett LC, Croarkin PE, Romanowicz M. 2022. Use of mobile and wearable artificial intelligence in child and adolescent psychiatry: Scoping review. *J Med Internet Res*. 24(3):e33560.
4. Arbgeli O, Rokach L, Cohen S. 2025. Wearable sensors for ensuring sports safety in children with autism spectrum disorder: A comprehensive review. *Sensors*. 25(5):1409.
5. Liu C, Yang F, Ge C, Shao Z. 2025. A dynamic precision evaluation system for physical education classroom teaching behaviors based on the CogVLM2-Video model. *Appl Sci*. 15(14):7712.
6. Oliveira C, Novelli F, Alves-Santos ET, Queiroz M, Ghezzi A, Cambri L. 2022. Physical activity influences heart rate variability in young adults, regardless of dextrose ingestion. *Eur J Appl Physiol*. 122(4):220–226.
7. Zhang J, Oh SS, Xu Y. 2026. Gamified physical education and cognitive performance among Chinese secondary school students: Cross-sectional moderation mediation study. *JMIR Serious Games*. 14:e81086.
8. Cao X. 2026. Multimodal sequence dynamics and convergence optimization in dual-stream LSTM networks for complex physiological state estimation. *Front Neurobot*. 20:1760494.
9. Fyntikakis E, Plakias S, Tsatalas T, Mina MA, Xenofondos A, Kokkoti C. 2026. Convergence of artificial intelligence and wearables in strength training and performance monitoring: A scoping review. *Appl Sci*. 16(7):3565.
10. Xu X. 2026. Multi-sensor fusion outperforms single indicators for fatigue prediction in university soccer players: A machine learning approach. *Front Physiol*. 17:1775906.
11. Alam S, Alam MAU. 2026. Explainable multitask burnout prediction using adaptive deep learning (EMBRACE) for resident physicians: Algorithm development and validation study. *JMIR AI*. 5:e57025.
12. Plaza-Florido A, Amaro-Gahete F, Acosta F, Sacha J, Alcantara J. 2022. Heart rate rather than heart rate variability is better associated with cardiorespiratory fitness in adults. *J Sports Sci*. 40(6):836–845.