



OPEN ACCESS

EDITED BY

Luiz Sanches Neto,
Federal University of Ceara, Brazil

REVIEWED BY

Abraham García-Fariña,
University of La Laguna, Spain
Omar Trabelsi,
University of Jendouba, Tunisia

*CORRESPONDENCE

TongKai Guan
✉ 870068818@qq.com

RECEIVED 17 November 2025

REVISED 08 February 2026

ACCEPTED 10 February 2026

PUBLISHED 23 March 2026

CITATION

Guan T, Chew RSY, Wen X and Huan B (2026) Using multimodal learning analytics as a formative assessment tool in AI-assisted physical education: a case study of Baduanjin teaching.

Front. Educ. 11:1744594.

doi: 10.3389/feduc.2026.1744594

COPYRIGHT

© 2026 Guan, Chew, Wen and Huan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Using multimodal learning analytics as a formative assessment tool in AI-assisted physical education: a case study of Baduanjin teaching

TongKai Guan^{1,2,3*}, Renee Shiun Yee Chew¹, XiaoMan Wen³ and BiXia Huan³

¹Faculty of Education and Liberal Arts, INTI International University, Nilai, Malaysia, ²Minbei Vocational and Technical College, Nanping, Fujian, China, ³Jimei University, Xiamen, China

Introduction: Traditional formative assessment in physical education (PE) often lacks objectivity and fails to capture the intricate multimodal dynamics of instructional behaviors. This study validates a Multimodal Learning Analytics (MMLA) framework to decode teaching effectiveness in a peer-coaching context using the traditional Chinese exercise, Baduanjin.

Methods: We analyzed synchronous data from 20 instructional dyads ($N = 20$), combining kinematic pose estimation (MediaPipe), speech fluency metrics (Whisper), and facial emotion recognition (OpenFace).

Results: Results from Spearman correlations and Mann-Whitney U tests revealed a “Precision-Economy” instructional archetype: high-performing instructors were characterized by superior kinematic fidelity ($r_s = .52$) and verbal economy ($r_s = .47$), rather than sheer feedback volume or emotional intensity. Counter-intuitively, excessive corrective feedback negatively correlated with learner skill gains ($r_s = -.41$), suggesting a cognitive load interference effect.

Discussion: These findings challenge the “more-is-better” pedagogical assumption and demonstrate that AI-driven analytics can objectively quantify the tacit mechanisms of embodied instruction, offering a scalable tool for teacher training and developing countries seeking to modernize PE assessment, while acknowledging the necessity for further validation in open-skill sports contexts.

KEYWORDS

embodied learning, formative assessment, Multimodal Learning Analytics (MMLA), physical education, teacher training and developing countries

1 Introduction

Physical education (PE) constitutes a critical domain for fostering psychomotor development and health literacy in adolescents (Slingerland et al., 2024). In the pedagogical sequence of PE, formative assessment serves as a pivotal mechanism for instructional refinement, providing pre-service teachers with the granular feedback necessary to bridge the gap between theoretical knowledge and embodied practice (Barrientos Hernán et al., 2023; Kormos, 2022). However, traditional assessment in PE teacher education (PETE)—primarily consisting of expert-led observations and qualitative rubrics—faces inherent challenges in the “embodied” classroom (Hay and Penney, 2013). The transient, multi-dimensional nature of instructional

behaviors, such as the micro-synchrony between a verbal cue and a motor demonstration, often eludes the naked eye, leading to evaluations that are characterized by high subjectivity and significant feedback latency (Barrientos Hernán et al., 2023; López-Pastor et al., 2013).

While video-based reflection is common, traditional observation often fails to objectively quantify the complex interplay between instructional modalities (López-Pastor et al., 2013; Barrientos Hernán et al., 2023; Stein et al., 2018). For instance, human observers struggle to measure the precise “signal-to-noise ratio”—such as the stability of a visual demonstration relative to the density of verbal cues (Blikstein and Worsley, 2016; Martínez Maldonado et al., 2018). Consequently, there is a need for objective, high-resolution diagnostic tools that can decode the structural patterns of embodied pedagogical expertise, moving beyond subjective rubrics to data-driven behavioral signatures (Blikstein, 2013; Cukurova et al., 2020).

Multimodal Learning Analytics (MMLA) has emerged as a promising frontier to address these limitations by synchronously capturing and interpreting diverse data streams, including instructional discourse, kinematic trajectories, and affective displays (Blikstein and Worsley, 2016). Although MMLA has been extensively applied in STEM and collaborative learning environments, its operationalization in physical education remains strikingly sparse (Wyant and Baek, 2019). Existing PE-related technology predominantly focuses on quantifying students’ physical activity levels (e.g., step count) rather than analyzing the qualitative nuances of the instructor’s multimodal triad—speech, movement, and emotion.

The present study addresses this empirical gap by utilizing MMLA as a formative assessment tool in a peer-coaching context. Using the traditional Chinese health Qigong, Baduanjin, as a case study, we recorded 20 instructional dyads ($N = 20$ instructors, $N = 20$ novices) through a multi-sensor array. Moving beyond simple activity tracking, we employ robust non-parametric statistical inference and micro-genetic qualitative analysis to identify the behavioral signatures of teaching effectiveness. Specifically, we investigate how the interplay of kinematic fidelity, speech fluency, and affective display collectively characterizes the instructional process. By examining whether teaching effectiveness is driven by a linear accumulation of these behaviors or by a specific “signal-to-noise” configuration (e.g., high precision with verbal economy), this research seeks to establish an evidence-based paradigm for PETE, shifting evaluation from holistic observation to data-driven diagnostic analytics.

2 Literature review

2.1 Multimodal learning analytics in embodied learning

Learning Analytics (LA) has traditionally focused on “digital traces” from online platforms, often ignoring the educationally significant interactions occurring in physical spaces (Blikstein, 2013). MMLA addresses this “street light effect” by integrating data from audio, video, and skeletal tracking to construct a holistic representation of learning (Su et al., 2020). In movement-intensive domains, MMLA offers a high-resolution “lens” to capture the micro-dynamic interactions that define expertise. Recent work has demonstrated that the integration of heterogeneous data types—such as

speech and gaze—can more accurately infer cognitive states than unimodal approaches (Martínez Maldonado et al., 2018; Zhao and Yu, 2024). However, translating these methods to the PE environment requires overcoming unique challenges in data synchronization and feature extraction (Liu, 2024), particularly for complex, coordinated motor tasks like Baduanjin.

2.2 The challenge of objectivity in formative PE assessment

Formative assessment, or “Assessment for Learning,” is essential for pre-service PE teachers to refine their pedagogical content knowledge (PETE) (Chng and Lund, 2018; Hay, 2006). Despite its importance, traditional observation methods suffer from inherent limitations. Inter-rater reliability is frequently compromised by the observers’ subjective stylistic preferences (Barrientos Hernán et al., 2023). Furthermore, standard rubrics typically lack the granularity to capture the degree of correctness in movement modeling; subtle deviations in joint angles can significantly alter the movement’s efficacy, yet these are often invisible in real-time (Wasik, 2011).

Critically, the timing and density of corrective feedback are major determinants of skill acquisition. According to the Guidance Hypothesis, while feedback is necessary, an excessive frequency of external information can hinder the learner’s ability to process task-intrinsic feedback, creating a “dependency effect” (Salmoni et al., 1984). Existing PETE frameworks lack the objective metrics to measure this delicate balance between instruction and silence.

2.3 Research gap and questions

Synthesizing the literature reveals a distinct gap: few empirical studies have simultaneously modeled the triad of PE teacher behavior—verbal, motor, and affective—within a unified analytical framework. While wearables have been used to track physical activity, we lack data on which specific combination of multimodal behaviors most effectively predicts motor skill acquisition (Wyant and Baek, 2019). This study aims to fill this gap by identifying the behavioral signatures of high-performing instructors using a mixed-methods MMLA approach. We pose two research questions:

RQ1: In a peer-coaching context, what behavioral signatures and patterns exist among the kinematic precision, speech fluency, and corrective feedback frequency of pre-service PE teachers?

RQ2: To what extent do these multimodal signatures distinguish high-performing from low-performing instructors, and how do they impact learner skill gain?

3 Methods

3.1 Research design

This study employed an Explanatory Mixed-Methods Design (Ivanova et al., 2006). Quantitatively, we captured high-dimensional multimodal data to identify statistical associations between instructional behaviors and teaching effectiveness (RQ2). Qualitatively, we

conducted a micro-genetic analysis of specific teaching episodes to contextualize the quantitative findings (RQ1).

3.2 Participants and dyadic composition

Participants were 40 students recruited from the second and third years of a Physical Education program at a sports vocational college (Age: $M = 20.4$, $SD = 1.2$). To ensure a clear knowledge gradient necessary for the peer-coaching model, participants were purposively assigned to 20 instructional dyads, each consisting of one Instructors (3rd year) had completed the “Traditional Chinese Sports’ module, whereas one Novice Learners (2nd year) were recruited prior to their enrollment in this module to ensure zero prior exposure to Baduanjin (Topping, 2005; Palinkas et al., 2015).

The Instructors ($n = 20$) were selected from the third years of a Physical Education program based on two criteria: (1) successful completion of foundational courses in sport pedagogy and exercise physiology, and (2) a high level of Pedagogical Content Knowledge (PCK) (Depaepe et al., 2013). To ensure baseline homogeneity in movement execution, all 20 instructors underwent a standardized 3-h intensive training session on the “Drawing the Bow to Shoot the Eagle” form, led by a national-level Qigong expert. Post-training assessments confirmed that all instructors achieved a minimum score of 8.5/10 (Chinese Health Qigong Association, 2011), ensuring that variations in teaching effectiveness were attributable to instructional delivery rather than the instructors’ own motor proficiency.

The Novice Learners ($n = 20$) were recruited from the second years of Physical Education program at a sports vocational college with zero prior experience in Baduanjin or related martial arts. A pre-test screening confirmed their baseline skill scores were uniformly low ($M = 1.1$, $SD = 0.3$), with no significant differences between the novices assigned to different dyads ($p > 0.05$). This strict differentiation established a controlled environment to measure the impact of instructional behaviors on genuine skill acquisition. Informed consent was obtained from all participants in accordance with the Declaration of Helsinki (World Medical Association, 2013).

3.3 Experimental task and standardization

The core task involved peer-coaching the “Drawing the Bow to Shoot the Eagle” (Second Form of Baduanjin) within a 15-min window. This specific movement was selected for its high pedagogical demand, requiring precise synchronized coordination of the upper and lower limbs, which elicits frequent multimodal interactions (verbal cues and visual modeling).

To maintain experimental control, the experiment was conducted in a quiet, well-lit university classroom, which was specifically arranged to ensure stable lighting conditions and a non-cluttered background for optimal video analysis (Druzhkov and Kustikova, 2016).

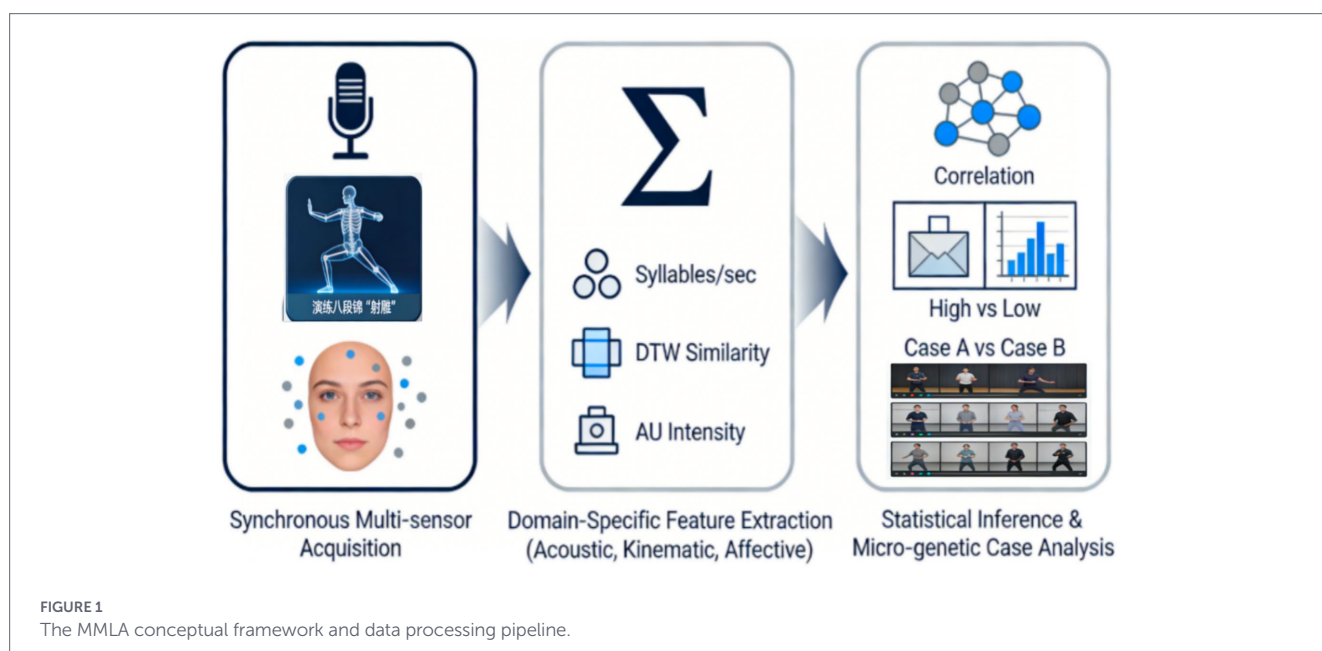
All instructors were provided with a standardized “Instructional Goal Card” to ensure they all aimed to teach the same three technical focal points (e.g., the “Horse Stance” depth, the “Bow Hand” tension, and the “Eye-Follow” coordination). This prevented the “Teaching Effectiveness” metric from being confounded by differences in the *content* of the instruction, thereby isolating the *process* variables (Speech Fluency, Pose Accuracy, etc.) as the primary objects of analysis (Newell, 1991; Schmidt and Lee, 2025).

Figure 1 illustrates the comprehensive research architecture, encompassing three integrated phases: (1) the synchronous acquisition of instructional discourse, kinematic demonstrations, and facial expressions; (2) the extraction of high-dimensional features through AI-based engines (Whisper, MediaPipe, and OpenFace); and (3) a dual-track analytical strategy that combines non-parametric statistical inference (Spearman correlation and Mann–Whitney U comparisons) with a micro-genetic qualitative analysis to elucidate the ‘precision-economy’ mechanisms of teaching effectiveness.

3.4 Data collection

Three categories of data were synchronously collected using a customized multi-sensor array:

Instructional discourse (audio): recorded using a Rode VideoMic NTG directional shotgun microphone mounted on a secondary tripod



at a distance of 3.5 meters. This setup utilized a super-cardioid polar pattern to selectively capture the instructor's vocal cues while minimizing ambient noise. Audio was sampled at 44.1 kHz, and transcriptions were generated via the OpenAI Whisper (Large-v3) engine, followed by manual verification to ensure a Word Error Rate (WER) below 3%. The transcribed text underwent word segmentation and part-of-speech tagging (Tian et al., 2020). Features extracted included the type of instruction (e.g., command, question, correction, encouragement) (Desai et al., 2020), keyword frequency, and speech rate (Edalatshams, 2022).

Movement demonstration (video): a high-definition (4 K) camera (Logitech Brio) recorded the instructor's full-body movements at 60 fps. To ensure alignment with our feature extraction pipeline, we utilized MediaPipe Pose (BlazePose GHUM model) to extract 33 skeletal landmarks in a 3D coordinate space. To mitigate potential jitter during complex transitions in the "Drawing the Bow" movement, a One Euro Filter was implemented for real-time temporal smoothing. The skeletal coordinate data output by MediaPipe were filtered (e.g., to reduce noise) and normalized (e.g., by body proportions) (Lin et al., 2023). Features characterizing the quality of the Baduanjin movement were extracted. These included: the accuracy of key joint angles (e.g., elbow angle during the "bow drawing" phase), based on skeletal joint angle estimation methods (Kim et al., 2023); the amplitude of movement (e.g., squat depth in the "horse stance"); postural symmetry; and the Dynamic Time Warping (DTW) similarity between the movement trajectory and a standard template.

Facial emotion (video): a secondary Logitech Brio 4 K camera focused on the instructor's face. OpenFace 2.0 (Baltrušaitis et al., 2018) was used to extract the intensity of 17 Facial Action Units (AUs). Features characterizing facial affect were extracted, including Action Units associated with both positive (e.g., happiness) and negative (e.g., frustration) emotional states (Baltrušaitis et al., 2018).

Figure 2A clearly illustrates the experimental setup, including the positions of the student, instructor, and cameras, as well as the real-time data capture and upload process. Figure 2B focuses on feature extraction, precisely annotating "Knee Joint Angle (e.g., horse stance Depth)," "Bow draw amplitude and symmetry," and the innovative "Core Stability (shaking trajectory)" on a MediaPipe skeletal frame of the "Drawing the Bow" movement.

3.5 Data analysis

3.5.1 Feature extraction and mathematical definitions

Speech Fluency (Syllables/s): To measure verbal processing speed accurately, we calculated the Articulation Rate excluding pause durations (De Jong and Wempe, 2009; Vercellotti, 2017; De Jong, 2018):

$$\text{Speech Fluency} = \frac{\text{Total Syllables Produced}}{\text{Phonation Time (seconds)}} \quad (1)$$

Pose Accuracy (Score): Instructor movements were compared against a Gold Standard Template generated from a National Health Qigong Champion (Professional Grade, >15 years experience). We employed Dynamic Time Warping (DTW) to calculate the inverse normalized distance between the instructor's skeletal trajectory and the expert template (higher score = greater similarity) (Zhao and Itti, 2018), normalized between 0 and 1.

Demonstration Variability: Defined as the coefficient of variation (CV) of the "Shoulder-Elbow-Wrist" angle during the holding phase.

Positive Emotion: Derived from the aggregated intensity of AU06 and AU12 (normalized 0-1) (Ekman, 1993; Prince et al., 2015).

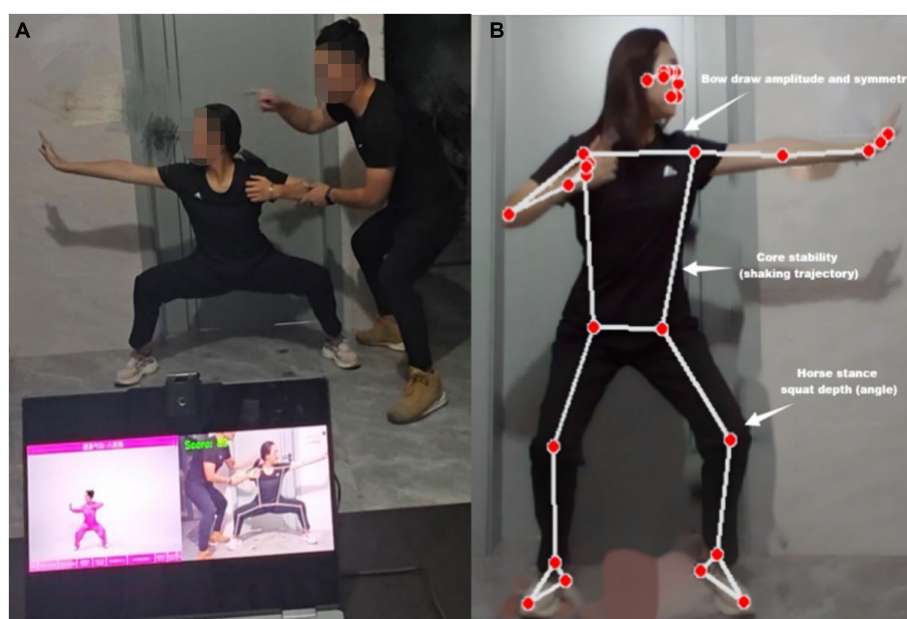


FIGURE 2

Schematic diagram of feature extraction from Baduanjin pose estimation. (A) Experimental setup of the peer-instruction scenario, illustrating the spatial arrangement of instructor and novice learner, camera positions, and the real-time multimodal data capture configuration. (B) Feature extraction interface based on MediaPipe skeletal tracking, highlighting key kinematic indicators including knee joint angle (horse stance depth), bow-draw amplitude and symmetry, and core stability (shaking trajectory).

Corrective Feedback Frequency: Defined as the total count of verbal interventions aimed at modifying the learner's motor execution (e.g., phrases including 'straighten,' 'lower,' 'correct'). This was identified through a rule-based keyword matching algorithm applied to the Whisper-generated transcripts.

Although a broad spectrum of linguistic (e.g., commands, encouragement) and affective (e.g., negative emotion) features were initially extracted, preliminary analysis indicated that Negative Emotion occurrences were negligible (floor effect, mean intensity < 0.05) due to the collaborative peer-coaching context. Similarly, instructional types such as Commands showed high collinearity with Corrective Feedback. Therefore, to ensure statistical power given the sample size and to focus on the study's core hypothesis regarding precision and economy, we retained only the most discriminative features: Speech Fluency, Corrective Feedback, Positive Emotion, Pose Accuracy, and Demonstration Variability for the final regression and correlation models.

3.5.2 Measurement of teaching effectiveness

Teaching Effectiveness was operationalized as a composite index ($Z_{composite}$) combining process and outcome:

Learner skill gain (outcome): we utilized the Pre-to-Post Gain Score ($Score_{post} - Score_{pre}$) to control for baseline differences. Skill assessment was based on official Health Qigong Competition Rules (0-10), evaluated by two blinded experts ($ICC = 0.88$). *T*-tests confirmed no significant difference in pre-test scores ($p > .05$).

Expert Rating of Instruction (Process): Two additional experts rated pedagogical behaviors using the adapted QualiTePE instrument (Herrmann et al., 2024).

The final dependent variable was the mean Z-score of the Skill Gain and the Expert Rating (hereafter referred to as "Teaching Effectiveness"). This composite approach ensures that the evaluation encompasses both the objective learning outcome and the pedagogical quality of the process.

3.5.3 Statistical modeling strategy

Given the sample size ($N = 20$), we prioritized robust non-parametric methods over complex multivariate regression to avoid overfitting (Field, 2024).

Spearman rank correlation (r_s): used to assess monotonic relationships between behavioral features and teaching effectiveness.

Group comparison (Mann-Whitney U): instructors were classified into "High Effectiveness" (Top 33%, $n = 7$) and "Low Effectiveness" (Bottom 33%, $n = 7$) groups to identify distinguishing behavioral characteristics.

Effect Size: Rank-biserial correlation (r_{rb}) was calculated for group comparisons.

4 Results

4.1 Descriptive statistics and correlations

To represent teaching effectiveness comprehensively, a composite Z-score ($Z_{composite}$) was calculated by averaging the standardized scores of Learner Skill Gain and Expert Ratings (as defined in Section 3.5.2). Descriptive analysis confirmed that the dependent variable, *Learner Skill Gain*, followed a normal distribution (Shapiro-Wilk, $p = 0.34$), with scores ranging from 1.5 to 5.5 ($M = 3.2, SD = 1.1$). This wide variance in learner outcomes provided a sufficient signal to investigate the impact of instructional heterogeneity. Table 1 details the Spearman rank correlations between the extracted multimodal features and teaching effectiveness.

As detailed in Table 1, the correlation analysis revealed distinct patterns across modalities. Visual modeling emerged as a primary driver of effectiveness, with *Pose Accuracy* demonstrating a strong positive association ($r_s = 0.52, p = 0.019$). This suggests that instructors who maintained high kinematic fidelity to the expert template significantly facilitated teaching effectiveness.

In the verbal domain, *Speech Fluency* also showed a moderate-to-strong positive correlation ($r_s = 0.47, p = 0.036$). Conversely, and counter-intuitively, *Corrective Feedback Frequency* exhibited a moderate negative correlation ($r_s = -0.41, p = 0.048$). This statistical trend implies a "less is more" dynamic, where excessive verbal interruptions may have approached a point of diminishing returns, potentially hindering the learners' processing of intrinsic feedback (Levine et al., 2019).

Crucially, the non-significant findings for Positive Emotion ($r_s = 0.19, p = 0.422$) offer a vital insight into the hierarchy of instructional behaviors. The weak correlation of Positive Emotion suggests that pedagogical affect is orthogonal to skill transmission in this specific motor task; while enthusiasm may foster engagement, the data indicates it cannot compensate for deficits in kinematic precision (as seen in the "High Emotion/Low Gain" outliers). Regarding kinematic stability, *Demonstration Variability* showed a negative trend ($r_s = -0.28, p = 0.232$), though it did not reach statistical significance (see Figure 3). This suggests that while lower variability (higher consistency) in visual modeling may theoretically benefit motor memory consolidation, the current sample size ($N = 20$) might be underpowered to detect this specific kinematic effect. However, the qualitative evidence from low-performing cases (see Section 4.3) suggests that

TABLE 1 Spearman rank correlations (r_s) between multimodal features and teaching effectiveness composite score ($N = 20$)

Variable	Mean (SD)	Effectiveness (r_s)	<i>p</i> -value
1. Pose Accuracy (Score)	0.78 (0.12)	0.52*	0.019
2. Speech Fluency (syll/sec)	3.25 (0.42)	0.47*	0.036
3. Demo Variability (CV %)	12.4 (3.1)	-0.28	0.232
4. Corrective Feedback (Freq)	21.3 (7.2)	-0.41*	0.048
5. Positive Emotion (Valence)	0.35 (0.18)	0.19	0.422

* $p < 0.05$.

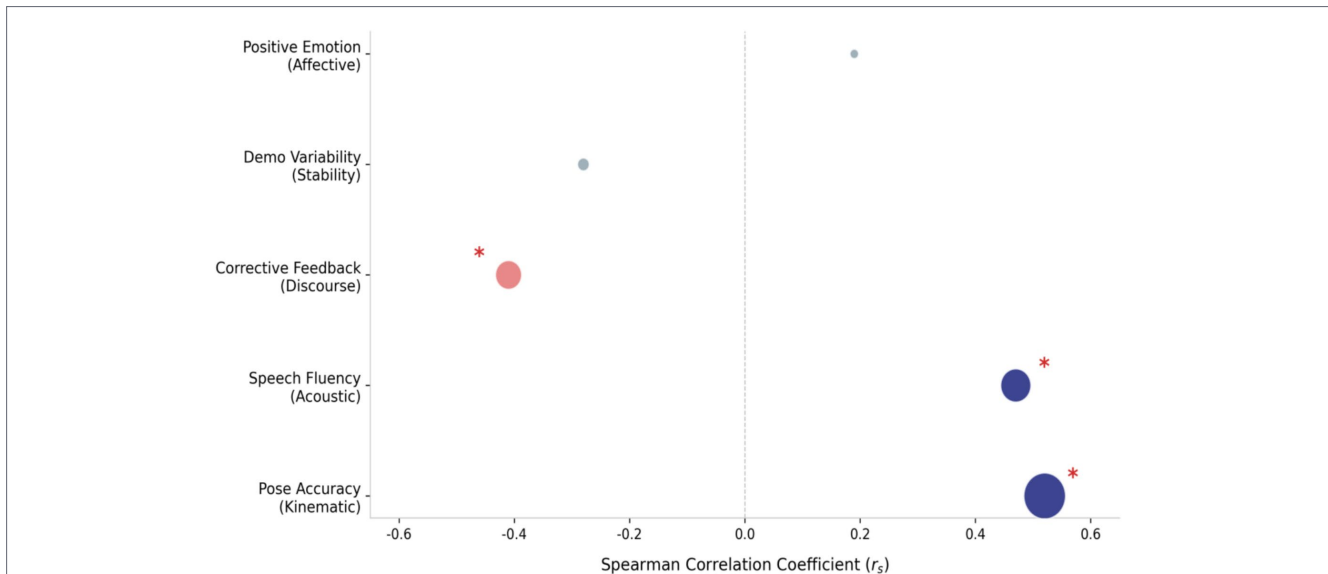


FIGURE 3 Integrated multi-modal correlation network. The distinct spatial clustering illustrates the "Precision-Economy" mechanism. Pose accuracy and speech fluency cluster as strong positive drivers (blue). Notably, positive emotion (gray) is isolated from the effectiveness axis, visualizing its role as a supportive rather than predictive factor in this psychomotor context. Corrective feedback (orange) pulls negatively, highlighting the risk of cognitive overload.

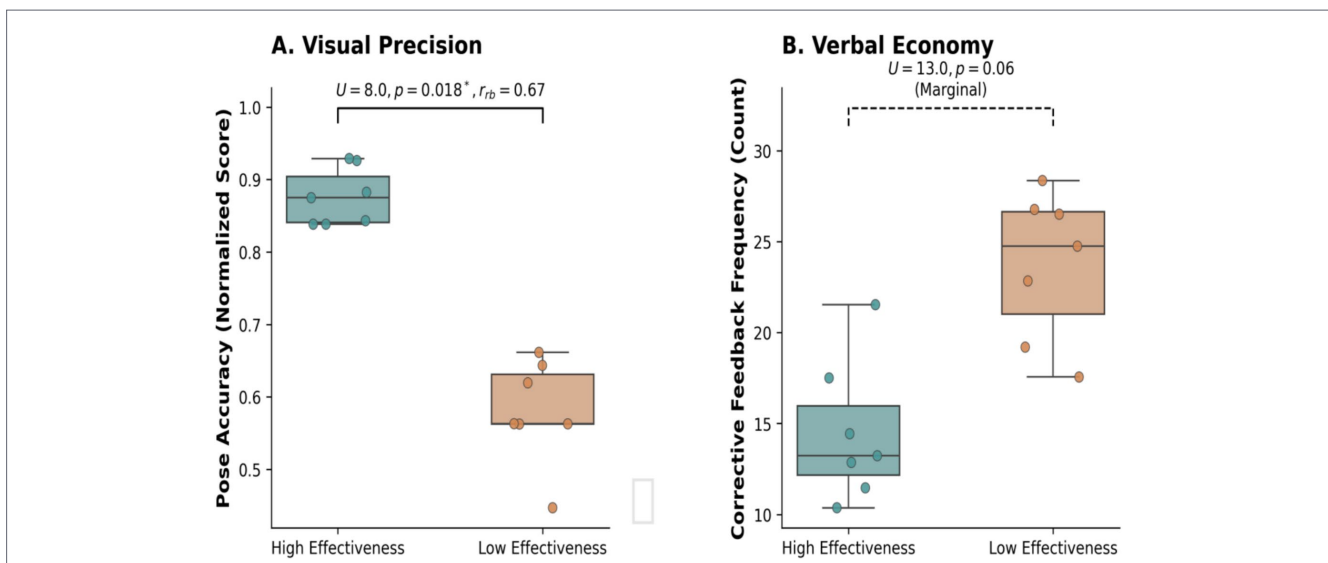


FIGURE 4 Comparative behavioral signatures of high vs. low effectiveness instructors. **(A)** Visual precision: High-effectiveness instructors demonstrate significantly greater pose accuracy compared to low-effectiveness instructors, indicating superior kinematic fidelity in movement modeling. **(B)** Verbal economy: High-effectiveness instructors provide fewer corrective feedback instances than low-effectiveness instructors, reflecting a more concise and strategically timed instructional approach.

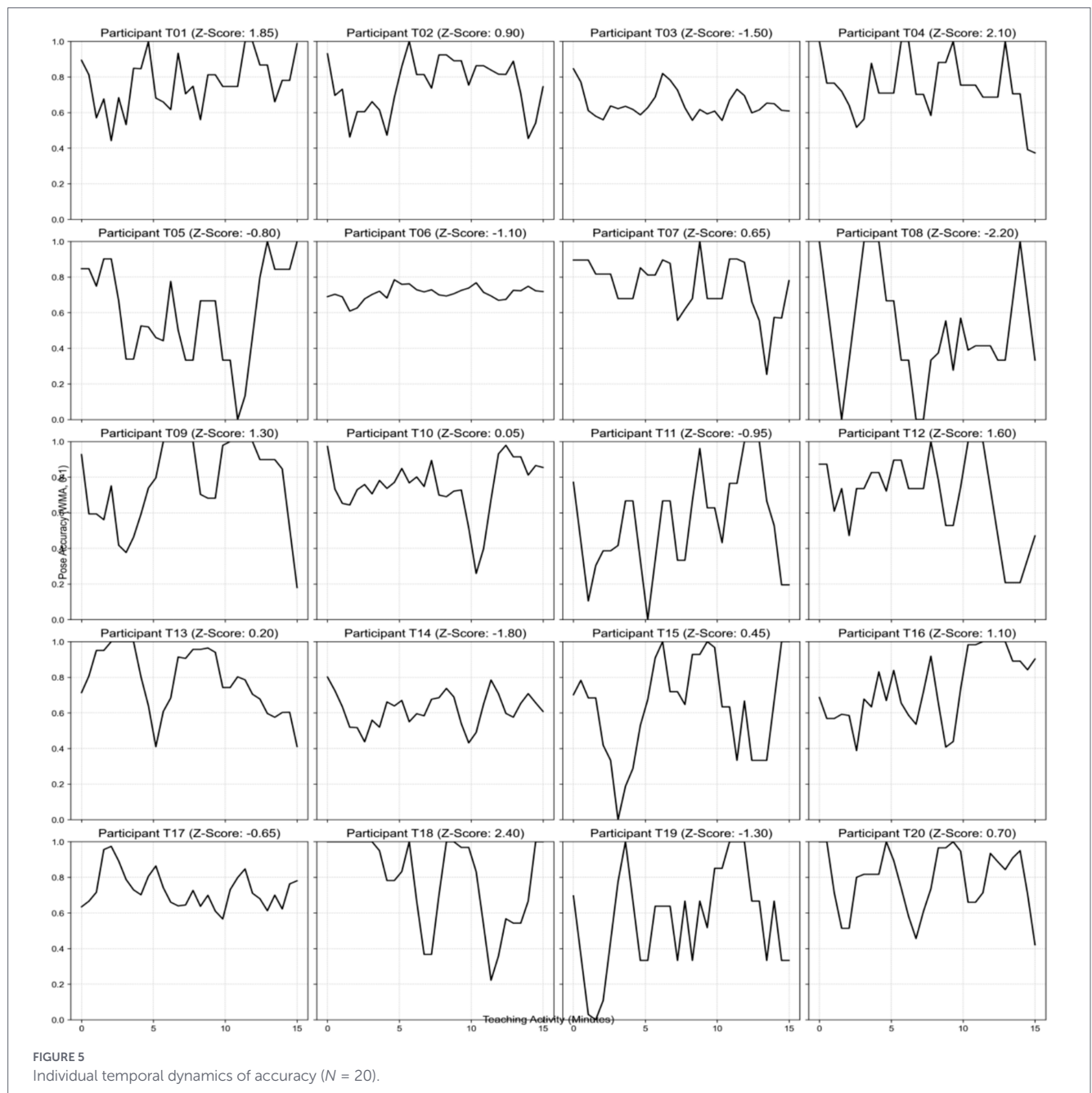
extreme variability ($CV > 15\%$) can indeed disrupt the learner’s observation process.

4.2 Comparative analysis of high vs. low effectiveness groups (RQ2)

To further distinguish the behavioral signatures of effective coaching, we stratified the sample into High Effectiveness ($n = 7, Z_{composite} = 4.8$) and Low Effectiveness ($n = 7, Z_{composite} = 1.9$) groups. Mann-Whitney U tests revealed significant distributional differences between these cohorts, as visualized in Figure 4.

Visual Precision: The High-performing group demonstrated significantly superior *Pose Accuracy* compared to their lower-performing counterparts ($U = 8.0, p = 0.018$). The rank-biserial correlation ($r_{rb} = 0.67$) indicates a large effect size, suggesting that the kinematic quality of the demonstration is a decisive factor distinguishing top-tier peer coaches (see Figure 4A).

Verbal Economy: A clear trend regarding verbal quantity was observed. High performers provided fewer total feedback instances ($M = 16.4$) compared to Low performers ($M = 26.1$). Although this difference was marginally significant ($U = 13.0, p = 0.06$), the separation in the interquartile ranges (see Figure 4B) suggests that effective



instructors prioritized concise, targeted cues over continuous commentary.

Emotional Display: Consistent with the correlation results, no significant difference was found in *Positive Emotion* intensity ($p > 0.05$), this null result is structurally significant: it indicates that high-performing instructors were not necessarily more enthusiastic or expressive than their lower-performing peers. Instead, effective peer-coaching was differentiated strictly by the “signal-to-noise ratio” of the instruction—high kinematic signal (Pose Accuracy) and low verbal noise (concise feedback)—rather than the affective valence of the delivery.

This inter-group difference was further validated in individual trajectories. Figure 5 displays the time series graph of Pose Accuracy for each of the 20 participants. This detailed view reveals the significant individual differences in instructional patterns that are averaged out in Figure 5. For instance, high-performing participants like T18 ($Z = 2.40$) show a high and relatively stable accuracy, while

low-performing participants like T08 ($Z = -2.20$) exhibit both low average accuracy and extreme volatility. This visualization makes the underlying data for our Demo_Variability (demonstration variability) feature explicit; one can observe how some participants (e.g., T06, $Z = -1.10$) show very little variability (low STDEV, “rigid”), others show moderate variability (e.g., T18, T12, “adaptive”), and some show chaotic variability (e.g., T08, T19, “chaotic”).

4.3 Qualitative micro-analysis of instructional episodes (RQ1)

To contextualize the quantitative findings and elucidate the mechanisms behind the “Verbal Economy” and “Visual Precision” effects, we conducted a micro-genetic analysis of two representative cases.

Case A (High Effectiveness, Instructor T18): Strategic Silence and Precision.

Instructor T18 ($Z_{gain} = +1.6$) exemplified a structured “Demonstrate-Observe-Feedback” cycle.

Behavioral Pattern: T18’s demonstrations were characterized by high stability (High Pose Accuracy). Crucially, T18 employed strategic post-response delays. After the novice attempted a movement, T18 typically paused for 3–5 s before intervening.

Instructional Consequence: This silence was not passive; video analysis showed T18 carefully observing the novice’s limb positioning. When feedback was finally delivered, it was concise and high-fluency (e.g., a single cue: “Sink your elbows, not your shoulders”), avoiding information overload. This pattern aligns with the quantitative finding that lower feedback frequency correlates with higher gains.

Case B (Low Effectiveness, Instructor T08): The “Enthusiastic Interference” Effect.

Instructor T08 ($Z_{gain} = -1.4$) presented a contrasting profile, characterized by high energy and high *Positive Emotion*.

Behavioral Pattern: T08 exhibited a pattern of concurrent interference. As shown in the audio-video timeline, T08 frequently shouted corrections during the novice’s motor execution (e.g., “Arms higher! Look at me! Now breathe!”).

Instructional Consequence: While well-intentioned, this high-frequency, overlapping feedback appeared to overwhelm the novice’s working memory. The novice was observed frequently freezing mid-movement to decode the verbal stream, disrupting motor flow. Additionally, T08’s tendency to speak while demonstrating resulted in higher Demonstration Variability ($CV = 15.8\%$). This multimodal dissonance—where high emotional intensity ($Valence = s0.8$) was coupled with unstable visual cues—created a “noisy” learning environment. This case explicitly validates the quantitative non-significance: high emotion could not rescue the learning outcome when visual fidelity was compromised.

Synthesizing the quantitative and qualitative evidence, the results suggest that teaching effectiveness in this context is not driven by the sheer quantity of feedback or emotional intensity. Instead, it is characterized by kinematic precision and verbal efficiency—specifically, the ability to provide accurate visual models coupled with concise, well-timed verbal cues.

5 Discussion

This study leveraged Multimodal Learning Analytics (MMLA) to decode the tacit behavioral dynamics of peer-coaching in Physical Education. By moving beyond descriptive statistics to a structural analysis of kinematic, linguistic, and affective streams, our findings propose a “Precision-Economy” instructional architecture. The data suggests that teaching effectiveness in closed-loop motor skills is not a cumulative function of “more behavior” (more feedback, more emotion), but rather a subtractive process of optimizing the signal-to-noise ratio—amplifying the visual model while actively suppressing verbal and kinematic interference.

5.1 The visual-verbal nexus: cognitive load and the “less-is-more” paradox

The prominent association between kinematic fidelity and skill acquisition, contrasted with the inverse relationship of feedback frequency, challenges the “content-heavy” tradition in pedagogy.

Theoretically, this affirms the primacy of the Mirror Neuron System (MNS) in motor learning, where the instructor’s body acts as the primary “biological blueprint” for the learner’s neural mapping (Rizzolatti and Craighero, 2004; Hardwick et al., 2018; Calvo-Merino et al., 2010).

However, the critical insight lies in the interference caused by excessive verbalization. Our findings extend the Guidance Hypothesis (Salmoni et al., 1984; Wulf and Lewthwaite, 2016) into the multimodal domain. While traditional theory posits that frequent feedback creates dependency, our multimodal lens reveals a more immediate cognitive bottleneck: the Split-Attention Effect (Mayer, 2002; Plass et al., 2010). When instructors deliver dense verbal corrections concurrently with movement (as observed in low-performing cases), they force the novice to engage distinct neural processors—phonological loops for speech and visuospatial sketchpads for movement—simultaneously. This dual-tasking likely overwhelms the learner’s limited working memory, leading to the “freezing” of degrees of freedom. Thus, “Speech Fluency” in this context should not be interpreted merely as “speaking speed,” but as processing efficiency—the ability to package complex motor commands into concise, high-density bursts that minimize the temporal tax on the learner’s attention.

5.2 Demonstration variability: functional adaptability vs. visual noise

The negative trend observed between demonstration variability and learning outcomes offers a nuanced perspective on Bernstein’s concept of “repetition without repetition.” While expert performers exhibit functional variability to adapt to environmental constraints (Davids et al., 2008; Ranganathan and Newell, 2013), our findings suggest that in a teaching context, variability functions differently.

For a novice learner attempting to construct a stable motor schema, inconsistencies in the instructor’s demonstration do not represent adaptability; they represent visual noise. According to Schema Theory (Schmidt, 1975; Schmidt and Lee, 2025), learners abstract a generalized motor program from observed examples. High variability in the model (e.g., fluctuating arm angles across repetitions) obscures the invariant topological features of the movement, forcing the learner to expend cognitive resources distinguishing “style” from “error.” Therefore, effective peer-coaching requires a “Standardized Template” strategy, where the instructor artificially suppresses their natural movement variability to provide a deterministic, noise-free visual reference (Scully and Carnegie, 1998; Scully and Newell, 1985; Ste-Marie et al., 2012).

5.3 The orthogonality of affect: distinguishing warmth from competence

Perhaps the most structurally revealing finding is the independence of Positive Emotion from immediate skill gain. This null result refines the application of Control-Value Theory (Robertson, 2015; Pekrun and Linnenbrink-Garcia, 2014) in physical education. It implies that pedagogical affect and technical transmission function on orthogonal axes.

While teacher enthusiasm is undisputedly vital for long-term motivation and psychological safety, our analysis indicates it is distinct from the mechanism of skill encoding. In the acute phase of motor learning, “warmth” cannot compensate for “kinematic imprecision.” Furthermore, high-arousal emotional displays, when not strictly coupled with instructional content, may act as seductive details (Harp and

Mayer, 1998; Wider, 2016; Park et al., 2015)—extrinsic stimuli that capture attention but divert processing resources away from the core motor task. This suggests a hierarchical model of effective coaching: emotional support acts as a foundational hygiene factor that permits learning to occur, but it is the precision of the “cold” cognitive mechanisms (visual modeling and verbal economy) that dictates the magnitude of skill acquisition.

5.4 Limitations and future directions

While this study establishes a baseline for multimodal assessment, limitations exist. The focus on a closed-loop skill (Baduanjin) limits generalization to open-skill sports where decision-making dynamics differ. Additionally, the cross-sectional design captures immediate acquisition rather than long-term retention. Future MMLA research should employ longitudinal designs to test whether the “verbal economy” that benefits novices remains effective as learners advance to autonomous stages of learning. Furthermore, integrating physiological sensors (e.g., fNIRS) could directly validate the cognitive load assumptions proposed in our “signal-to-noise” framework.

6 Conclusion

This study demonstrates the transformative potential of Multimodal Learning Analytics (MMLA) in deciphering the latent behavioral mechanisms of effective Physical Education. By systematically integrating kinematic, linguistic, and affective data streams, we have moved beyond the subjectivity of traditional observation to identify a “Precision-Economy” instructional archetype. Our findings reveal that superior teaching effectiveness in closed-loop motor skills is not predicated on the accumulation of pedagogical behaviors (more feedback, higher enthusiasm), but rather on the optimization of the signal-to-noise ratio: providing a high-fidelity visual blueprint (Pose Accuracy) while adhering to “strategic verbal silence” (Verbal Economy) to preserve the learner’s limited cognitive bandwidth.

Theoretically, this research extends the Guidance Hypothesis and Cognitive Load Theory into the realm of embodied pedagogy, establishing that pedagogical affect and technical transmission operate on orthogonal axes. While positive emotion provides the necessary psychological safety, it cannot compensate for the “visual noise” introduced by kinematic variability or the “cognitive interference” caused by excessive feedback. Practically, the automated pipeline developed herein offers a scalable prototype for AI-assisted formative assessment. It suggests that future teacher education should pivot from training general “instructional volume” to cultivating “multimodal efficiency”—empowering educators to recognize when to speak, how to demonstrate with deterministic stability, and crucially, when to remain silent to facilitate the learner’s organic self-organization.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements as the study involved routine classroom-based educational activities without any medical, clinical, or biomedical interventions. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

TG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. RC: Supervision, Writing – review & editing. XW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing. BH: Visualization, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by INTI International University, Nilai, Malaysia, which provided reimbursement for publication-related fees (article processing charges).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Artificial intelligence–based tools were used solely for the visualization of conceptual figures. All scientific content, analyses, interpretations, and conclusions were developed and written by the authors.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Baltrušaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). "OpenFace 2.0: Facial behavior analysis toolkit," in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (59–66). IEEE.
- Barrientos Hernán, E. J., López-Pastor, V. M., Lorente-Catalán, E., and Kirk, D. (2023). Challenges with using formative and authentic assessment in physical education teaching from experienced teachers' perspectives. *Curr. Stud. Health Phys. Educ.* 14, 109–126. doi: 10.1080/25742981.2022.2060118
- Blikstein, P. (2013). "Multimodal learning analytics," in Proceedings of the Third International Conference on Learning Analytics and Knowledge (102–106).
- Blikstein, P., and Worsley, M. (2016). Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. *J. Learn. Anal.* 3, 220–238. doi: 10.18608/jla.2016.32.14
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2010). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex* 15, 1243–1249. doi: 10.1093/cercor/bhi007
- Chinese Health Qigong Association (2011). Health Qigong competition rules International Health Qigong Federation. Available online at: https://www.ihqfo.org/uploadfile/file/20180124/20180124033712_48891.pdf
- Chng, L. S., and Lund, J. (2018). Assessment for learning in physical education: the what, why and how. *J. Phys. Educ. Recreat. Dance* 89, 29–34. doi: 10.1080/07303084.2018.1503119
- Cukurova, M., Giannakos, M., and Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *Br. J. Educ. Technol.* 51, 1441–1449. doi: 10.1111/bjet.13015
- David, K., Button, C., and Bennett, S. (2008). *Dynamics of skill acquisition: A constraints-led approach*. New York, NY: Human Kinetics.
- De Jong, N. H. (2018). Fluency in second language testing: insights from different disciplines. *Lang. Assess. Q.* 15, 237–254. doi: 10.1080/15434303.2018.1477780
- De Jong, N. H., and Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Speech Commun.* 51, 385–396. doi: 10.1016/j.specom.2008.05.002
- Depaepe, F., Verschaffel, L., and Kelchtermans, G. (2013). Pedagogical content knowledge: a systematic review of the way in which the concept has pervaded mathematics educational research. *Educ. Res. Rev.* 10, 12–28. doi: 10.1016/j.tate.2013.03.001
- Desai, T., Dakle, P., and Moldovan, D. (2020). "Joint learning of syntactic features helps discourse segmentation," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (1073–1080).
- Druzhkov, P. N., and Kustikova, V. D. (2016). A survey of deep learning methods and software tools for image classification and object detection. *Pattern Recognit. Image Anal.* 26, 9–15. doi: 10.1134/S1054661816010065
- Edalatihams, S. (2022). *A prosodic corpus of teaching assistant classroom speech: Discourse intonation and information structure [Doctoral dissertation, Iowa State University]*. Ames, IA: Iowa State University Digital Repository.
- Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 384–392. doi: 10.1037/0003-066X.48.4.384
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. 6th Edn. London: Sage.
- Hardwick, R. M., Caspers, S., Eickhoff, S. B., and Swinnen, S. P. (2018). Neural correlates of action: comparing meta-analyses of imagery, observation, and execution. *Neurosci. Biobehav. Rev.* 94, 31–44. doi: 10.1016/j.neubiorev.2018.08.003
- Harp, S. F., and Mayer, R. E. (1998). How seductive details do their damage: a theory of cognitive interest in science learning. *J. Educ. Psychol.* 90, 414–434. doi: 10.1037/0022-0663.90.3.414
- Hay, P. (2006). "Assessment for learning in physical education" in *Handbook of physical education*. eds. D. Kirk, D. Macdonald and M. O'Sullivan (London: SAGE Publications), 312–325.
- Hay, P., and Penney, D. (2013). *Assessment in physical education: A socio-cultural perspective*. London: Routledge.
- Herrmann, C., Crapa, A., Langer, W., Adamakis, M., Borghouts, L., Cools, W., et al. (2024). *The QualiTePE instrument to evaluate quality of teaching in physical education: Documentation of items and scales in English, German, French, Italian, Spanish, Dutch, Swedish, Slovenian, Czech and Greek*. London: Zenodo.
- Ivankova, N. V., Creswell, J. W., and Stick, S. L. (2006). Using mixed-methods sequential explanatory design: from theory to practice. *Field Methods* 18, 3–20. doi: 10.1177/1525822X05282260
- Kim, J.-W., Choi, J.-Y., Ha, E.-J., and Choi, J.-H. (2023). Human pose estimation using MediaPipe Pose and optimization method based on a humanoid model. *Appl. Sci.* 13:2700. doi: 10.3390/app13042700
- Kormos, E. (2022). A comparison of preservice teacher perceptions of instructor video and text-based feedback. *SN Soc. Sci.* 2:153. doi: 10.1007/s43545-022-00413-9
- Levine, D., Cheng, A., Olaleye, D., Leonardo, K., Shifrin, M., and Ishii, H. (2019). "AUFLIP – An auditory feedback system towards implicit learning of advanced motor skills," in Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (1–6). Association for Computing Machinery.
- Lin, Y., Jiao, X., and Zhao, L. (2023). Detection of 3D human posture based on improved MediaPipe. *J. Comput. Commun.* 11, 102–121. doi: 10.4236/jcc.2023.112008
- Liu, G. (2024). Multimodal analysis and optimisation strategy of teaching behaviour in physical education classroom. *Appl. Math. Nonlinear Sci.* 9. doi: 10.2478/amns-2024-1506
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., and Macdonald, D. (2013). Alternative assessment in physical education: a review of international literature. *Sport Educ. Soc.* 18, 57–76. doi: 10.1080/13573322.2012.713860
- Martínez Maldonado, R., Echeverría, V., Santos, O., Dias Pereira Santos, A., and Yacef, K. (2018). "Physical learning analytics: A multimodal perspective," in Proceedings of the 8th International Conference on Learning Analytics and Knowledge (375–379). Association for Computing Machinery.
- Mayer, R. E. (2002). "Multimedia learning" in *Psychology of learning and motivation*. eds. K. W. Spence and J. T. Spence, vol. 41 (New York, NY: Academic Press), 85–139.
- Newell, K. M. (1991). Motor skill acquisition. *Annu. Rev. Psychol.* 42, 213–237. doi: 10.1146/annurev.ps.42.020191.001241
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., and Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Admin. Policy Mental Health Mental Health Serv. Res.* 42, 533–544. doi: 10.1007/s10488-013-0528-y
- Park, B., Flowerday, T., and Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Comput. Hum. Behav.* 44, 267–278. doi: 10.1016/j.chb.2014.10.061
- Pekrun, R., and Linnenbrink-Garcia, L. (2014). *International handbook of emotions in education*. London: Routledge.
- Plass, J. L., Moreno, R., and Brünken, R. (2010). *Cognitive load theory*. Cham: Springer.
- Prince, E. B., Martin, K. B., Messinger, D. S., and Allen, M. (2015). "Facial action coding system" in *Environmental psychology & nonverbal behavior*. eds. P. Ekman, W. V. Friesen and J. Hager (Cham: Springer), 1–15.
- Ranganathan, R., and Newell, K. M. (2013). Changing up the routine: Intervention-induced variability in motor learning. *Exercise Sport Sci. Rev.* 41, 64–70. doi: 10.1097/JES.0b013e318259beb5
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Robertson, L. (2015). *International handbook of emotions in education*. London: Routledge.
- Salmoni, A. W., Schmidt, R. A., and Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychol. Bull.* 95, 355–386. doi: 10.1037/0033-2909.95.3.355
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychol. Rev.* 82, 225–260. doi: 10.1037/h0076770
- Schmidt, R. A., and Lee, T. D. (2025). *Motor learning and performance: From principles to application*. 6th Edn. New York, NY: Human Kinetics.
- Scully, D., and Carnegie, E. (1998). Observational learning in motor skill acquisition: a look at demonstrations. *Ir. J. Psychol.* 19, 472–485. doi: 10.1080/03033910.1998.10558208
- Scully, D. M., and Newell, K. M. (1985). Observational learning and the acquisition of motor skills: toward a visual perception perspective. *J. Hum. Mov. Stud.* 11, 169–186. doi: 10.1016/S0167-9457(85)80006-0
- Slingerland, M., Weeldenburg, G., and Borghouts, L. (2024). Formative assessment in physical education: teachers' experiences when designing and implementing formative assessment activities. *Eur. Phys. Educ. Rev.* 30, 620–637. doi: 10.1177/1356336X241237398
- Stein, M., Janetzko, H., Lamprecht, A., Breitkreutz, T., Zimmermann, P., Goldlücke, B., et al. (2018). Bring it to the pitch: combining video and movement data to enhance team sport analysis. *IEEE Trans. Vis. Comput. Graph.* 24, 13–22. doi: 10.1109/TVCG.2017.2745181
- Ste-Marie, D. M., Law, B., Rymal, A. M., Jenny, O., Hall, C., and McCullagh, P. (2012). Observation interventions for motor skill learning and performance: an applied model for the use of observation. *Int. Rev. Sport Exerc. Psychol.* 5, 145–176. doi: 10.1080/1750984X.2012.665076
- Su, M., Cui, M., and Huang, X. (2020). Multimodal data fusion in learning analytics: a systematic review. *Sensors* 20:6856. doi: 10.3390/s20236856
- Tian, Y., Song, Y., and Xia, F. (2020). "Joint Chinese word segmentation and part-of-speech tagging via multi-channel attention of character N-grams," in Proceedings of COLING 2020: The 28th International Conference on Computational Linguistics (2073–2084). International Committee on Computational Linguistics.
- Topping, K. J. (2005). Trends in peer learning. *Educ. Psychol.* 25, 631–645. doi: 10.1080/01443410500345172
- Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: a longitudinal study. *Appl. Linguist.* 38, 90–111.
- Wasik, J. (2011). Kinematic analysis of the side kick in Taekwon-do. *Acta Bioeng. Biomech.* 13:4.

- Wider, W. (2016). The effect of occupational stress social support as moderator. *Australian Journal of basic and applied sciences* 10, 54–65.
- World Medical Association (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053
- Wulf, G., and Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin. Rev.* 23, 1382–1414. doi: 10.3758/s13423-015-0999-9
- Wyant, J. D., and Baek, J. H. (2019). Re-thinking technology adoption in physical education. *Curr. Stud. Health Phys. Educ.* 10, 3–17. doi: 10.1080/25742981.2018.1552499
- Zhao, J., and Itti, L. (2018). ShapeDTW: shape dynamic time warping. *Pattern Recognit.* 74, 171–184. doi: 10.1016/j.patcog.2017.09.020
- Zhao, Y., and Yu, B. (2024). Construction of a classification model for teacher and student behavior in physical education classrooms – based on multimodal data. *Appl. Math. Nonlinear Sci.* 9. doi: 10.2478/amns-2024-1818