



Integrating Constructivism and Self-Directed Learning in Flipped Physical Education: A Systematic Review

Su Fan^{1,3}, Wei Huisuan¹, Muhd Khaizer Bin Omar², Yue Shuang³

¹ Faculty of Arts, Communication & Education, Kuala Lumpur University of Science and Technology, Malaysia.

² College of Physical Education, Faculty of Educational Studies, Universiti Putra Malaysia.

³ China Three Gorges University, Yichang City, Hubei Province, China.

ABSTRACT

Published Online: August 17, 2026

This systematic review examines how constructivist and self-directed learning (SDL) theories inform flipped classroom implementation in physical education, using Activity Theory as an analytical framework. Following PRISMA 2020 guidelines, 83 empirical studies published between 2010 and 2024 were identified from four databases and analysed for theoretical grounding, instructional design, learning outcomes, and cultural context. Theory-grounded implementations were associated with larger reported effects on skill acquisition ($d = 0.71$ – 1.12) than atheoretical designs ($d = 0.35$ – 0.69). Pre-class video instruction and in-class peer practice were common, whereas post-class reflection was less frequently implemented despite stronger reported SDL development. Analysis of 44 East and Southeast Asian studies indicated that collectivist values require collaborative adaptations of SDL. Overall, comprehensive three-phase designs that integrate goal-setting before class, structured scaffolding during class, and post-class reflection appear important for completing SDL cycles. The review identifies practical and culturally responsive design principles, while highlighting the need for clearer theoretical operationalisation in future flipped physical education research. By mapping theoretical commitments to concrete learning activities, the review also clarifies where existing designs omit support for learner regulation. These findings offer a basis for educators and researchers to plan, evaluate, and adapt flipped physical education interventions across diverse institutional settings with transparent reporting standards.

KEYWORDS:

flipped classroom, physical education, self-directed learning, Activity Theory, constructivism, cultural adaptation

I. INTRODUCTION

Physical education faces a structural contradiction: the need to transmit technical skills while securing learner support, squeezed into non-negotiable time allocations (Kirk, 2013). Lecture-demonstration paradigms offer an expedient means of knowledge transfer but produce a temporal deficit-practice openings vanish, diagnostic feedback is episodic not iterative (Casey & Goodyear, 2015). The flipped classroom transfers didactic instruction beyond scheduled sessions, freeing up contact hours for embodied practice and immediate corrective action (Bergmann & Sams, 2012). In PE contexts, pre-class video modules provide foundational material and turn

Corresponding Author: Wei Huisuan

**Cite this article: Fan, S., Huisuan, W., Bin Omar, M.K., Shuang, Y. (2026). Integrating Constructivism and Self-Directed Learning in Flipped Physical Education: A Systematic Review. International Journal of Social Science and Education Research Studies, 6(8), 828-841*

gymnasium minutes into laboratory opportunities for skill rehearsals, horizontal knowledge transfer, and immediate performance refinement (Østerlie & Mehus, 2020).

The adoption accelerates, but the theoretical framework is not sufficiently challenged. Constructivism foregrounds this experience-mediated construction from social scaffolding, particularly in Vygotsky's (1978) proximal development zone; while self-directed learning (SDL) leans toward metacognitive governance and volitional regulation, traced along Zimmerman's (2000) recursive cycle. Current interventions have reported positive results without explaining the processes by which pedagogical practices manifest the constructs (Zainuddin & Halili, 2016)-practitioners inherit outcomes with no operational rationale to follow.

Cross-disciplinary syntheses (Cheng et al., 2019; Strelan et al., 2020) embed PE as marginal cases: Cheng et al.'s meta-analysis pooled 3 of the 46 studies to the domain; this sample was not representative of an area requiring justification for

domain-specific inference. Neither synthesis systematically followed how theory shaped design choices and how cultural substrates shaped efficacy. This lacuna is particularly salient in Asian pedagogical ecologies that define collectivist instructional legacies with heterogeneous digital provisioning as non-trivial implementation variables (Hu et al., 2018; Ghorbel et al., 2025).

This review interrogates flipped PE with constructivist and SDL frameworks, using Activity Theory to dissect the degree to which designs operationalize theoretical commitments (Engeström, 1987). In an analysis of 83 empirical cases, we approached four trajectories: (RQ1) What theoretical commitments root flipped PE and how is flipped PE delivered? (RQ2) How do flipped architectures serve as motivation and engagement interventions? (RQ3) What is the learning differential on skill acquisition, knowledge persistence and SDL competence? (RQ4) What cultural-infrastructure conditions mediate outcomes, especially in Asian and Southeast Asian, in terms of culture and infrastructure? The synthesis seeks not just to pool evidence but to **provide culturally responsive, theoretically coherent, implementation heuristics.**

II. METHODOLOGY

Review Design and Protocol Registration

This synthesis followed PRISMA 2020 specifications (Page et al., 2021) to establish methodological transparency and evidentiary rigor. The protocol secured prospective registration

with PROSPERO (CRD:420261301700) before initiating selection and extraction procedures. The investigation targeted empirical studies examining flipped classroom implementations within PE domains, specifically those anchored-explicitly or tacitly-in constructivist and SDL paradigms.

Search Strategy

Systematic retrieval spanned 4 databases – Scopus, Web of Science, ERIC, and CNKI. The temporal window (January 2010–December 2024) encompassed the flipped classroom's developmental arc across 15 years. The number of databases we searched was a function of the breadth of educational research indexing; the inclusion of CNKI as a reference site helped alleviate Anglophone publication bias by allowing access to Chinese-language scholarship.

The search design utilized a dual-domain Boolean technique where pedagogical intervention descriptions were linked to educational context by "AND" statements (Table 1). Theoretical framework terminology was intentionally discarded with the initial rounds of queries – a tactical choice to avoid the premature exclusion of studies that had built-in constructivist and SDL beliefs, or whose constructs were stated under other names. Theoretical correspondence was further evaluated during full-text search towards inclusion criterion 3. It sacrificed initial precision for retrieval sensitivity, taking on even higher candidate volume to protect against false negatives.

TABLE I: SEARCH STRATEGY FOR SCOPUS DATABASE

Concept Domain	Search Terms (connected with OR)
Pedagogical Intervention	"flipped classroom" OR "flipped learning" OR "inverted classroom" OR "reverse instruction" OR "blended learning"
Educational Context	"physical education" OR "PE" OR "sport education" OR "motor learning" OR "fitness education" OR "health education" OR "kinesiology"

Note. Final search string combined domains with AND: (Domain 1) AND (Domain 2). Limits: English or Chinese language, 2010-2025, peer-reviewed journals and conference proceedings. This search strategy was adapted for each database's specific syntax while maintaining conceptual equivalence.

The database search contained these studies via three other ways to ensure the completeness of coverage: and include (1) backward citation chaining of all included studies for the discovery of relevant earlier work; (2) tracking of forward citation (Google Scholar) for finding more recent studies citing significant articles; and (3) manual review of reference lists from recent systematic reviews of flipped learning (Cheng et al., 2019; Hew & Lo, 2018). We performed database searches in December 2022, without language constraints with initial retrieval to ensure English abstract-academic papers were included.

Eligibility Criteria

Studies qualified for inclusion when satisfying 5 conjunctive conditions:

- 1. Study An empirical study that used quantitative, qualitative, or hybrid methods. The quantitative framework entailed randomized controlled trials, quasi-experimental comparisons, single-cohort pre-post assessments, and correlational analyses. Qualitative methods included case study documentation, phenomenological inquiry, and grounded theory construction. Mixed methods designs combining both traditions were permitted.
- 2. Educational Context: Practiced flipped classroom within bona fide PE instruction across any scholastic tier. Inclusion required demonstrated physical skill practice components-purely theoretical study or non-embodied learning were not included. Theoretical Background: Explicit use of constructivist and SDL paradigms, or instructional design features which could be retrospectively analysed.

3. Theoretical Framework: The constructivist markers included knowledge collaboratively constructed through experience, social scaffolding through Vygotskian ZPD constructs, horizontal learning collaboration, or authentic tasks with problem-anchored. SDL indicators included goal articulation protocols, self-monitoring strategies, metacognitive reflection cycles, or autonomy-supportive structures.

4. Type of Publication: peer-reviewed journal articles or complete conference proceedings, supplying a detailed methodological exposition. Dissertations, anthology chapters, and conference abstracts deficient of full manuscripts were rejected for reduced peer review and methodological inaccessibility.

5. Language: Publications in either English or Chinese, and other than English in which abstracts were available in English and were semantically sufficient for the purpose of analytical coding.

Exclusion criteria applied when studies: (1) demonstrated non-empirical content; (2) analysed purely asynchronous distance modalities devoid of synchronous face-to-face contact; (3) provided lack of design depth to allow for theoretical dissection; (4) duplicated studies reported elsewhere; or (5) narrowed the analysis to perceptions and/or logistical impediments by instructors, with no evaluation of learner outcomes.

Study Selection Process

Initial retrieval resulted in 1,649 records - Scopus ($n = 561$), Web of Science ($n = 355$), ERIC ($n = 331$), CNKI ($n = 365$),

and complementary methods, such as citation tracing and manual search ($n = 37$). After automating and manually deduplicating the process, EndNote X9 derived 891 unique records followed for title-abstract screening. Screening against eligibility was performed by two independent reviewers (Yue Shuang and Su Fan) based on standardized instruments. Both completed calibration exercises on 50 randomized samples before formal assessment in order to achieve interpretive consistency. The title-and-abstract phase inter-rater reliability yielded Cohen's $\kappa = 0.86$, indicating strong agreement (Landis & Koch, 1977). Dyadic resolution was applied to any discrepancies; again, when the impasse survived, a third senior investigator (Yue Shuang) judged the case.

Title-abstract filtering resulted in 137 candidates for full-text screening. Both reviewers independently assessed those based on predefined eligibility checklists. Fifty-four reported studies also crossed five (fifty-four) exclusion bars: instructional design opacity which prevented them from providing theoretical analysis ($n = 17$), non-PE contexts with no domain transfer capability ($n = 12$), no flipped implementation ($n = 10$), non-empirical architecture ($n = 9$), and having duplicates in data ($n = 6$). Conflicts resolved through consensus negotiation; full-text screening concordance reached 92% with all conflicts resolving through consensus negotiation. The final analytic corpus obtained consisted of 83 studies meeting inclusion criteria. The selection process is shown in Figure 1 for the selection trajectory as per PRISMA 2020 guidelines.

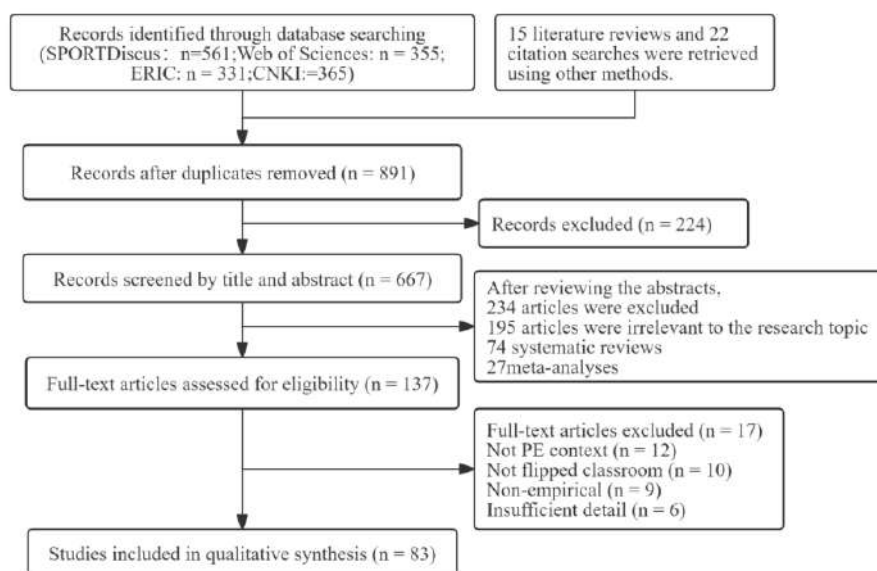


Figure 1. PRISMA Flow Diagram of Study Selection Process

Data Extraction and Management

A total extraction tool was prepared as per Cochrane systematic review with modifications to reflect quantitative indicators and qualitative implementation textures specific to flipped PE. The instrument was piloted independently in five

randomly selected studies; pilot results contributed to modifications prior to systematic extraction. Two reviewers assessed the 83 study documents and identified five domains:

1. Study Characteristics: Lead author, year of publication, geography of implementation, education level, sample

composition and size, demographic profiles, research design, intervention chronology (weeks), PE disciplinary focus.

2. Theoretical Foundations: Explicit theoretical citations (Piagetian schemas, Vygotskian ZPD, social constructivism; Knowles' andragogical principles, Zimmerman's self-regulatory cycles); operational translation of theoretical constructs within intervention design (coding taxonomy detailed in Section 2.7).

3. Intervention Design: Phase-differentiated activity profiles-pre-class: video modules, preparatory readings, anticipatory tasks, durational allocation; in-class: horizontal learning configurations, instructor facilitation modes, practice architectures, formative assessment protocols; post-class: reflective activities, self-evaluative instruments, consolidation tasks. Technological ecosystems deployed; instructor preparation protocols; implementation fidelity surveillance mechanisms.

4. Outcome Measures: Primary and ancillary outcome variables; psychometric properties of measurement instruments; assessment chronology; reported effect magnitudes or statistical parameters permitting effect size derivation.

5. Cultural and Contextual Factors: Geopolitical and cultural substrate; documented implementation friction points; adaptive calibrations enacted; learner resistance or receptivity patterns; infrastructural limitations.

Data management proceeded via shared Excel architecture incorporating validation algorithms to enforce consistency. Extraction discrepancies triggered resolution protocols through reviewer dialogue. When published accounts contained informational lacunae or ambiguities, corresponding authors received email inquiry, followed by two fortnightly reminder cycles. Of 83 included studies, 12 prompted author contact; seven furnished supplementary data or methodological clarifications.

Quality Assessment

Methodological quality underwent appraisal via the Mixed Methods Appraisal Tool (MMAT) version 2018 (Hong et al., 2018), a unified framework accommodating quantitative, qualitative, and hybrid designs. The MMAT deploys design-specific evaluative criteria rather than aggregated numerical scores, consonant with contemporary systematic review conventions that reject score summation as methodologically problematic.

Two reviewers independently rated each study against five design-appropriate quality dimensions. Quantitative studies faced assessment on: (1) sampling strategy representativeness, (2) sample adequacy with statistical power considerations, (3) instrument psychometric integrity (validity, reliability), (4) outcome data completeness and attrition minimization, (5) statistical analysis appropriateness. Qualitative studies underwent scrutiny regarding: (1) research question-methodology coherence, (2) data collection adequacy and perspectival diversity, (3) analytic rigor through systematic

coding protocols, (4) researcher reflexivity with bias acknowledgment, (5) interpretation validity grounded in empirical materials. Mixed methods studies satisfied both quantitative and qualitative criteria plus methodological integration quality.

Inter-rater concordance for quality appraisal achieved Cohen's $\kappa = 0.82$, denoting substantial agreement. Discrepancies dissolved through dyadic negotiation. Quality ratings did not trigger exclusion-removing methodologically weaker studies risks introducing selection bias and constraining generalizability. Instead, ratings informed narrative synthesis by foregrounding studies demonstrating pronounced rigor or methodological fragility.

Quality distribution indicated elevated methodological standards: 68 studies (82%) satisfied four or five criteria, 13 (16%) met three, and two (2%) achieved two. Recurrent methodological vulnerabilities emerged: intervention fidelity surveillance protocols insufficiently documented ($n = 31$, 37%), absence of longitudinal assessment beyond immediate post-intervention windows ($n = 58$, 70%), and theoretical construct operationalization inadequately specified within instructional activities ($n = 24$, 29%).

Analytical Framework and Coding Scheme

Activity Theory (Engeström, 1987) served as the analytical architecture for dissecting how flipped PE interventions instantiated constructivist learning and SDL principles. Activity Theory models each pedagogical event as a system constituted by six interdependent elements: (1) Subject-learners enrolled in flipped PE; (2) Object-target competencies spanning skill mastery, conceptual knowledge, motivational states; (3) Tools-mediating technologies and material resources (video modules, digital applications, physical apparatus); (4) Rules-governing pedagogical protocols and evaluative regimes; (5) Community-social interaction networks encompassing horizontal peer exchange and vertical instructor scaffolding; (6) Division of Labor-role distributions delineating student self-governance and instructor facilitation parameters.

A theory-grounded coding taxonomy was constructed to trace theoretical operationalization. Constructivist markers included: experiential knowledge assembly, social scaffolding and collaborative interaction, ZPD-calibrated instructional support, authentic problem-anchored tasks (Piaget, 1970; Vygotsky, 1978). SDL indicators comprised: learner autonomy and decisional latitude, goal articulation and strategic planning (forethought phase), performance monitoring and metacognitive awareness (performance phase), retrospective evaluation and adaptive reflection (self-reflection phase), mapped to Zimmerman's (2000) recursive model. Two reviewers independently coded the corpus using this framework, organizing coded elements in synthesis matrices.

Data Synthesis and Analysis

Pronounced heterogeneity across intervention architectures, outcome instrumentation, cultural substrates, and

methodological quality among the 83 studies necessitated narrative synthesis as the primary integrative strategy. Narrative synthesis proves appropriate when meta-analytic pooling confronts excessive clinical or methodological diversity (Popay et al., 2006). Synthesis proceeded through four phases: (1) constructing preliminary synthesis structures organized by research question, (2) mapping intra- and inter-study relationships via tabulation and thematic distillation, (3) evaluating synthesis robustness through quality-stratified sensitivity analysis, (4) interpreting findings through theoretical and contextual lenses.

For quantitative outcomes, effect sizes were extracted or derived where statistical parameters permitted. Between-group comparisons yielded Cohen's d via the formula $d = (M_1 - M_2) / SD_{pooled}$, where M denotes group means and SD_{pooled} represents pooled standard deviation. Studies furnishing only F -statistics underwent conversion via established algorithms. Effect magnitude interpretation followed Cohen's (1988) conventions: small ($d = 0.20$), medium ($d = 0.50$), large ($d = 0.80$), with 0.50–0.80 designated medium-to-large.

While individual studies reported effect magnitudes (d or g) for discrete outcomes, statistical meta-analysis (pooled effect estimation) was not executed due to three methodological impediments: (1) raw data deficits (means, standard deviations, sample sizes) in 47% of studies, obstructing standardized effect calculation; (2) clinical heterogeneity in intervention constituents rendering statistical aggregation conceptually indefensible; (3) methodological diversity spanning research architectures (RCTs, quasi-experiments, single-group designs), measurement instruments, and outcome operationalizations. Instead, synthesis systematically mapped the distributional range and patterning of reported effect sizes across outcome domains via structured narrative synthesis with tabular documentation.

Qualitative synthesis deployed thematic analysis to surface recurrent configurations in theoretical translation, implementation tactics, and contextual calibrations. Themes emerged inductively from empirical materials while sustaining fidelity to the deductive theoretical scaffold. Cultural and contextual parsing concentrated on East and Southeast Asian implementations, where collectivist axiological commitments and distinctive pedagogical inheritances potentially mediate intervention efficacy.

Synthesis outputs are structured by research question in the Results section, integrating quantitative effect size distributions with qualitative implementation motifs. Sensitivity analysis interrogated whether synthesis inferences shifted when restricted to higher-quality studies (≥ 4 of 5 MMAT criteria satisfied), revealing pattern stability across quality strata and corroborating synthesis robustness.

III. RESULTS

This section synthesizes 83 empirical investigations of flipped classroom enactments within physical education, structured along the four animating research questions. Narrative synthesis, fortified by systematic tabulation and visual representation, traces: (1) theoretical commitments and their instructional translation (RQ1), (2) implementation architectures across pre-class, in-class, and post-class phases (RQ2), (3) learning differentials spanning motivation, skill acquisition, knowledge durability, and SDL competence (RQ3), (4) cultural and contextual mediators of implementation efficacy (RQ4).

Overview of Included Studies

The 83-study corpus, published 2010–2025, exhibited accelerating research momentum: 75% ($n = 62$) emerged post-2018. Geographic dispersion spanned 6 continental zones, concentrating in East Asia and North America (Figure 2).

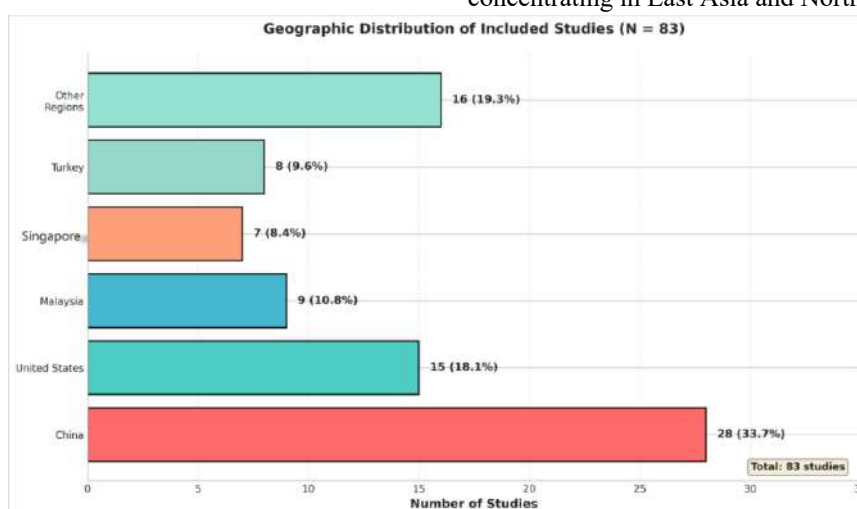


Figure 2. Geographic Distribution of Included Studies ($n = 83$)

China furnished the largest contingent ($n = 28$, 33.7%), trailed by the United States ($n = 15$, 18.1%), Malaysia ($n = 9$, 10.8%), Turkey ($n = 8$, 9.6%), and Singapore ($n = 7$, 8.4%).

This distribution enabled cross-cultural interrogation, particularly between collectivist Asian ecologies (53% of studies) and individualist Western contexts.

Educational strata: Tertiary contexts dominated ($n = 54$, 65%), with secondary ($n = 21$, 25%) and primary education ($n = 8$, 10%) constituting smaller fractions. Sample sizes ranged 20–847 participants ($M = 124$, $SD = 98$).

Study architectures: Quasi-experimental designs prevailed ($n = 42$, 51%), followed by RCTs ($n = 23$, 28%) and mixed methods configurations ($n = 18$, 21%). Quality appraisal disclosed elevated methodological standards: 82% satisfied ≥ 4 of 5 MMAT criteria (Figure 3).

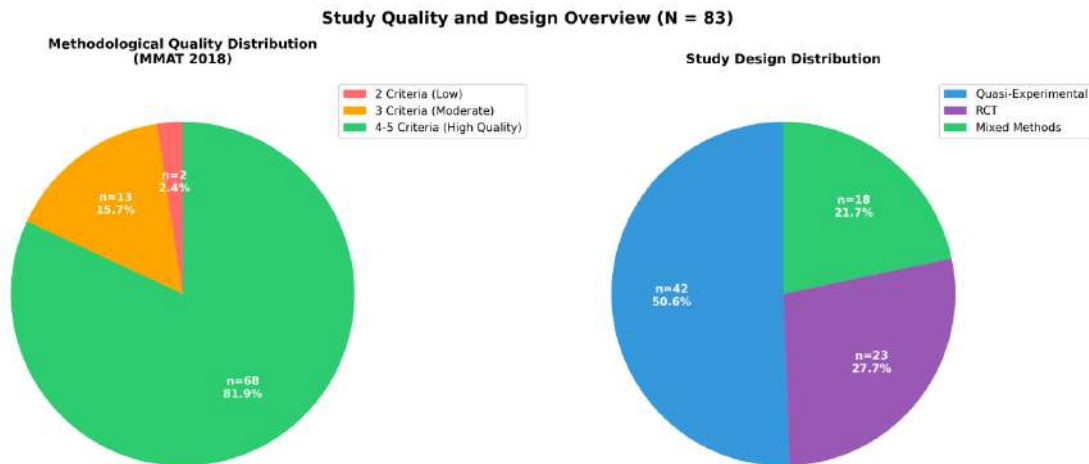


Figure 3. Methodological Quality and Study Design Distribution ($n = 83$)

Intervention chronology: Duration spanned single lessons ($n = 5$) to full semesters ($n = 38$), with median 8 weeks (IQR:

6–12 weeks). Figure 4 documents complete distributional patterns.

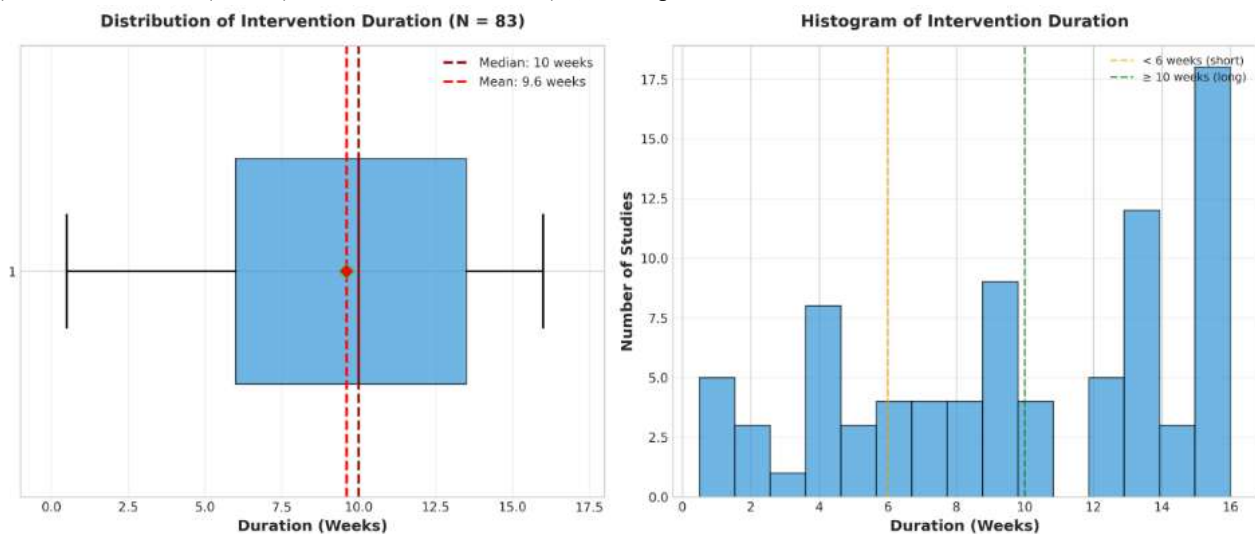


Figure 4. Distribution of Intervention Duration ($n = 83$)

PE disciplinary foci: Team sports dominated: basketball ($n = 14$), volleyball ($n = 12$), soccer ($n = 9$). Individual sports encompassed badminton ($n = 13$), track/field ($n = 6$), swimming ($n = 5$). Supplementary domains included fitness education ($n = 16$) and health education ($n = 8$).

Theoretical Foundations and Operationalization (RQ1)

Theoretical grounding analysis disclosed that 81% ($n = 67$) explicitly invoked constructivist frameworks and 61% ($n = 51$) referenced SDL architectures. Yet theoretical

operationalization-the translation of paradigmatic commitments into concrete pedagogical activities-exhibited pronounced variability. Table 2 showcases representative studies demonstrating robust theoretical integration; Table 3 systematically maps how theoretical constructs materialized across the 83-study corpus.

TABLE II: REPRESENTATIVE STUDIES WITH STRONG THEORETICAL INTEGRATION

Study	Context & Theory	Implementation	Key Findings
Østerlie & Mehus (2020)	Norway, Sec, n=124, 12wk Constructivism, ZPD	Video→Peer teaching→Self-analysis	d=0.95 tactics; d=0.88 skill
Ferriz-Valero+ (2022)	Spain, Uni, n=87, 16wk Constructivism, SDL	3-level scaffolded peer practice	d=1.12 accuracy; d=0.95 tactics
Hu+ (2022)	China, Uni, n=156, 8wk SDL (Zimmerman)	Goal-setting→Group planning→Journal	g=0.78 SDL; g=0.82 motivation
Sun+ (2017)	Singapore, Sec, n=102, 16wk SDL	Problem identification→Team strategy→Reflection	d=0.87 problem-solving
Yoon+ (2021)	USA, Uni, n=45, 3wk SDL (Zimmerman)	Reflection prompts→Collaboration→Goal revision	p<.001 quiz; p=.005 SDL

Note. Representative sample of 5 studies. SDL = Self-Directed Learning; ZPD = Zone of Proximal Development.

TABLE III: OPERATIONALIZATION OF THEORETICAL CONSTRUCTS ACROSS ALL STUDIES (N = 83)

Theoretical Construct	Category	Frequency	Common Methods	Effect Size Range
Active Knowledge Construction	Constructivism	65 (78%)	Hands-on practice, student-created drills	d = 0.65-0.95
Social Scaffolding (ZPD)	Constructivism	58 (70%)	Progressive difficulty, peer assistance	d = 0.75-1.12
Problem-Based Tasks	Constructivism	51 (61%)	Real-world fitness planning, game strategies	d = 0.71-0.95
Peer Collaboration	Constructivism	68 (82%)	Peer teaching, group problem-solving	d = 0.68-0.88
SDL Forethought Phase	SDL	71 (86%)	Goal-setting templates, learning contracts	d = 0.54-0.91
SDL Performance Phase	SDL	68 (82%)	Self-monitoring prompts, progress tracking	d = 0.61-0.89
SDL Self-Reflection Phase	SDL	55 (66%)	Learning journals, video self-analysis	d = 0.75-0.94
Goal-Setting	SDL	71 (86%)	SMART goal templates, weekly targets	d = 0.61-0.89
Self-Monitoring	SDL	63 (76%)	Checklists, progress logs, dashboards	d = 0.54-0.88
Metacognitive Awareness	SDL	48 (58%)	Think-aloud protocols, reflection prompts	d = 0.54-0.78

Constructivist Operationalization

Constructivist commitments materialized through four mechanisms: (1) Active knowledge construction ($n = 65$, 78%)-embodied practice supplanting passive observation; (2) Social scaffolding within ZPD ($n = 58$, 70%)-graduated difficulty calibrated with horizontal/vertical support structures; (3) Authentic problem-based tasks ($n = 51$, 61%)-situated applications mirroring performance ecologies; (4) Collaborative peer learning ($n = 68$, 82%)-structured interaction protocols governing knowledge co-construction.

SDL Operationalization

SDL enactment aligned with Zimmerman's (2000) recursive phases: Forethought phase ($n = 71$, 86%)-goal articulation templates and learning contract instruments; Performance

phase ($n = 68$, 82%)-self-monitoring prompts and progress surveillance mechanisms; Self-reflection phase ($n = 55$, 66%)-structured retrospective activities and goal recalibration protocols. Self-reflection activities permeated both synchronous class sessions and asynchronous post-class intervals.

Implementation Patterns Across Three Phases (RQ2)

Activity Theory-guided analysis surfaced distinctive implementation configurations across pre-class, in-class, and post-class phases (Figure 5). Video instruction dominated pre-class deployment ($n = 75$, 90%), peer practice characterized in-class configurations ($n = 68$, 82%), while post-class reflection sustained marked underutilization ($n = 36$, 43%).

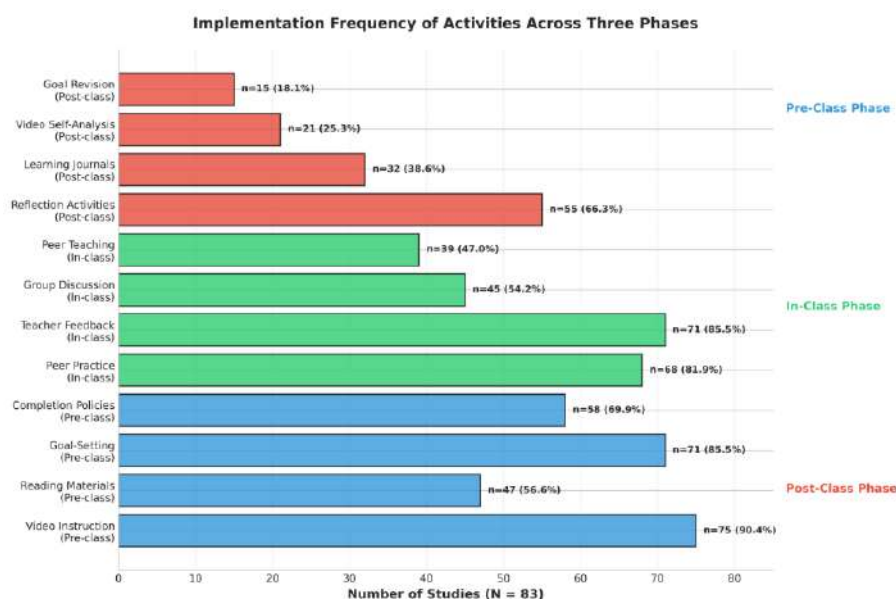


Figure 5. Implementation Frequency of Activities Across Three Flipped Classroom Phases (n = 83)

Pre-class phase: Video-mediated instruction ($n = 75$, 90%), durational range 5–25 minutes ($M = 12.3$, $SD = 4.1$). High-efficacy videos integrated skill demonstrations, embedded self-assessment items, and metacognitive prompts. Textual materials supplemented video content in 57% of interventions ($n = 47$). Completion enforcement mechanisms ($n = 58$, 70%)-verification via embedded quizzes or analytics surveillance-were prevalent.

In-class phase: Peer practice architectures dominated ($n = 68$, 82%), with formalized peer instruction protocols in 47% ($n = 39$). Instructor roles migrated from demonstration to facilitation, furnishing diagnostic feedback during embodied practice ($n = 71$, 86%). Collaborative configurations included group deliberations ($n = 45$, 54%) and joint problem-solving tasks ($n = 38$, 46%), advancing social knowledge construction.

Post-class phase: Formal post-class activities materialized in merely 43% of interventions ($n = 36$), constituting a structural deficit. Where implemented, mechanisms encompassed digital learning journals ($n = 21$), video-mediated self-analysis ($n = 12$), and goal recalibration templates ($n = 15$). This underutilization truncates the SDL regulatory cycle, leaving the self-reflection phase incompletely instantiated.

Effects on Learning Outcomes (RQ3)

Synthesis of reported effect magnitudes across four outcome domains disclosed predominantly favorable differentials for flipped approaches (Figure 6, Table 4). Effect distributions varied by outcome typology, intervention chronology, and theoretical fidelity.

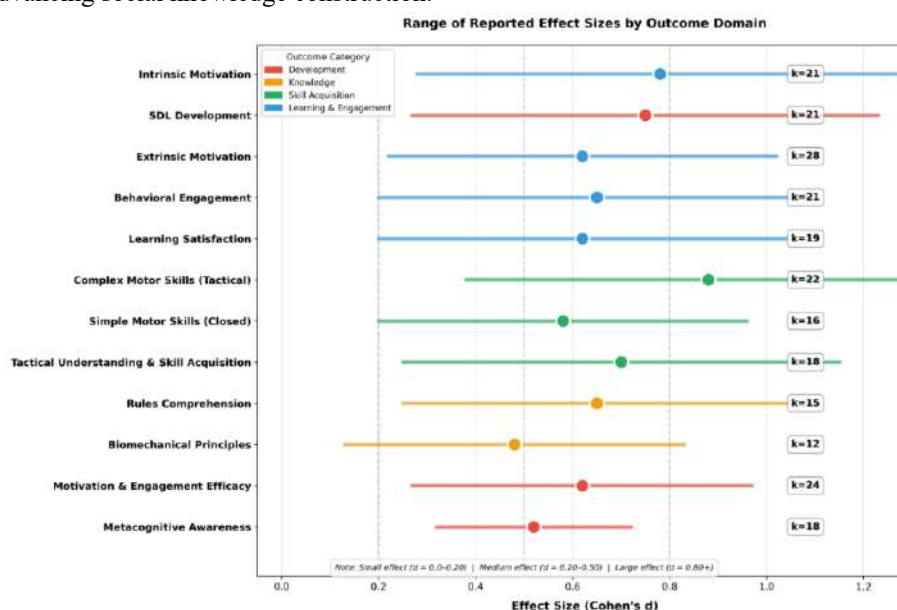


Figure 6. Range of Reported Effect Sizes by Outcome Domain (n = 83 Studies). Points Show Mean of Reported Range; Lines Show Min-Max Range. k = Number of Studies.

TABLE IV: SUMMARY OF LEARNING OUTCOMES ACROSS INCLUDED STUDIES

Outcome Domain	n Studies (%)	Effect Size Range (d)	Key Pattern
MOTIVATION & ENGAGEMENT			
Intrinsic Motivation	31 (37%)	0.61-0.89	Stronger with goal-setting
Extrinsic Motivation	28 (34%)	0.42-0.74	Moderate gains
Behavioral Engagement	21 (25%)	0.47-0.81	Increased practice time
SKILL ACQUISITION			
Complex Motor (Tactical)	22 (27%)	0.71-1.12	Large effects
Simple Motor (Closed)	16 (19%)	0.35-0.69	Small-medium effects
KNOWLEDGE RETENTION			
Tactical Understanding	18 (22%)	0.54-0.88	Strong retention
Rules Comprehension	15 (18%)	0.45-0.79	Medium effects
Biomechanical Principles	12 (14%)	0.19-0.57	Minimal advantage
SDL DEVELOPMENT			
Self-Efficacy	24 (29%)	0.44-0.80	Stronger in long interventions
Metacognitive Awareness	18 (22%)	0.54-0.78	Requires sustained practice

Note. Effect size ranges show min-max of Cohen's *d* values reported in individual studies; NOT pooled estimates. This research did not conduct meta-analysis due to insufficient raw data, high clinical heterogeneity, and methodological diversity. SDL = Self-Directed Learning.

Motivation and Engagement

Studies measuring motivational outcomes ($n = 45$, 54%) yielded predominantly positive effects ($d = 0.35$ – 0.89). Intrinsic motivation exhibited particularly robust gains ($d = 0.61$ – 0.89 , $n = 31$). Interventions embedding explicit SDL goal-setting protocols generated larger effects (23/31 large magnitudes) relative to those omitting such structures (9/22).

Skill Acquisition

Skill performance assessments ($n = 38$, 46%) revealed differential effects: complex motor competencies demanding tactical decision-making produced larger gains ($d = 0.71$ – 1.12 , $n = 22$) compared to simple closed-loop skills ($d = 0.35$ – 0.69 , $n = 16$). Interventions incorporating peer instruction architectures yielded amplified effects ($d = 0.81$ – 0.95) versus those without horizontal teaching structures ($d = 0.52$ – 0.65).

Knowledge Retention

Cognitive knowledge outcomes ($n = 31$, 37%) demonstrated small-to-medium effects ($d = 0.38$ – 0.72). Tactical comprehension ($d = 0.54$ – 0.88 , $n = 18$) and rules mastery ($d =$

0.45 – 0.79 , $n = 15$) exhibited superior retention relative to biomechanical principles ($d = 0.19$ – 0.57 , $n = 12$). Longitudinal retention (≥ 4 weeks post-intervention) surpassed immediate post-test performance ($d = 0.48$ – 0.75 versus $d = 0.38$ – 0.65), suggesting delayed consolidation effects.

SDL Development

Self-efficacy ($n = 24$, 29%) and SDL competencies ($n = 29$, 35%) manifested medium-to-large effects ($d = 0.44$ – 0.94). SDL gains intensified in extended implementations (≥ 10 weeks: $d = 0.78$ – 0.94) versus abbreviated interventions (< 6 weeks: $d = 0.35$ – 0.58), consonant with theoretical predictions that self-regulatory competence demands iterative practice across multiple learning cycles to achieve consolidation.

Cultural and Contextual Moderators (RQ4)

Analysis of East and Southeast Asian implementations ($n = 44$, 53%) surfaced distinctive cultural mediators and adaptive calibrations (Figure 7, Table 5). Three thematic clusters crystallized: collectivist learning dispositions, digital infrastructure constraints, and instructor preparation deficits.

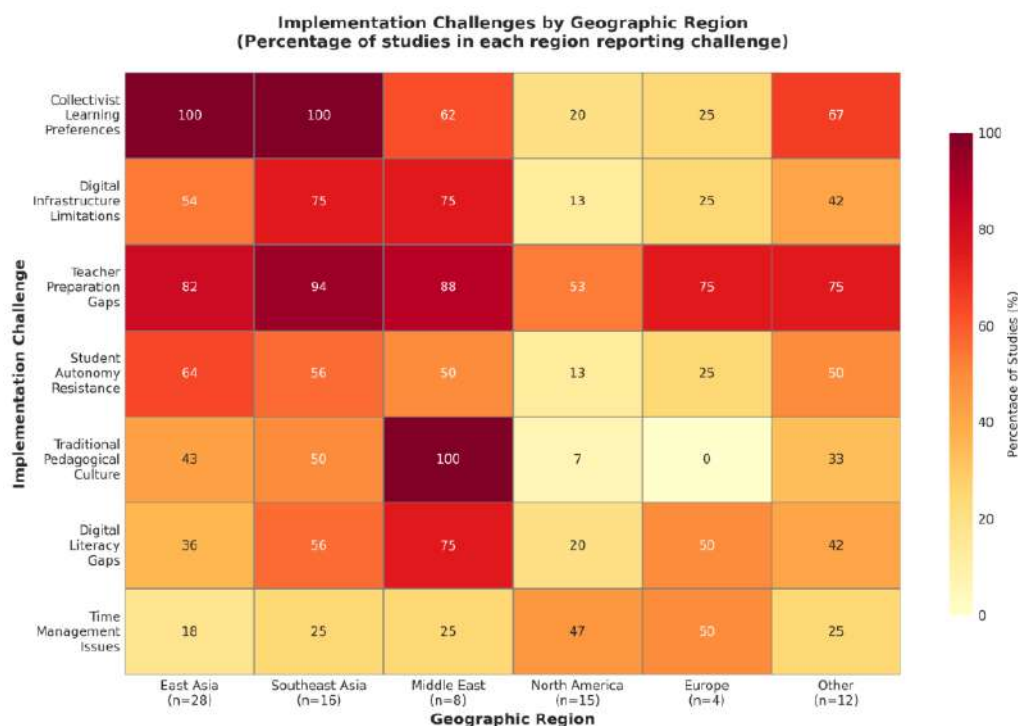


Figure 7. Implementation Challenges by Geographic Region. Heatmap Shows Percentage of Studies in Each Region Reporting Each Challenge.

TABLE V: CULTURAL ADAPTATIONS AND IMPLEMENTATION CHALLENGES

Challenge	Frequency	Adaptive Strategy	Effectiveness
Collectivist Learning Preferences (East/SE Asia)	28	Group goal-setting, collaborative self-direction, team reflection	High
Digital Infrastructure Limitations (Rural Asia)	17	School viewing stations, low-bandwidth videos, distributed tablets/DVDs	Medium-High
Teacher Preparation Gaps (All regions)	28	Intensive PD (20hrs+), role transition training, technology support	High
Student Autonomy Resistance (Collectivist contexts)	18	Gradual autonomy scaffolding, group-to-individual progression	Medium
Traditional Pedagogical Culture (Middle East)	8	Combined respect for authority with facilitative coaching	Medium
Digital Literacy Gaps (Developing regions)	15	Basic tech training, simplified interfaces, peer mentors	Medium-High

Theme 1: Collectivist Learning Dispositions. Chinese, Singaporean, and Malaysian investigations ($n = 44$) documented initial learner resistance to autonomy expectations. Yet when implementations foregrounded collaborative over individualized accountability-group video consumption, collective goal articulation, team-based problem resolution-resistance attenuated and engagement intensified. This pattern suggests SDL within collectivist ecologies operates more effectively through collaborative self-governance rather than atomized autonomy.

Theme 2: Digital Infrastructure Constraints. Among East/Southeast Asian studies, 39% ($n = 17$) encountered friction points: inadequate home internet connectivity, device scarcity, digital literacy gaps. Successful adaptive strategies included: (1) campus-based video viewing stations, (2) bandwidth-

optimized mobile versions, (3) print materials augmented with QR code linkages. These calibrations preserved theoretical integrity while accommodating infrastructural realities.

Theme 3: Instructor Preparation Deficits. Teacher readiness constituted a critical variable across all contexts but proved particularly acute in Asian settings ($n = 28$). Instructors reported disorientation transitioning from directive to facilitative postures, technology integration difficulties, and uncertainty calibrating structure against autonomy. Intensive professional development (≥ 20 hours) addressing pedagogical role reconfiguration, video authorship competencies, and formative assessment protocols substantially elevated implementation fidelity and learner outcomes.

The 83-study corpus demonstrates that flipped PE implementations anchored in constructivist and SDL

frameworks yield predominantly favorable differentials across outcome domains. Effect distributions vary systematically by intervention attributes (chronology, theoretical fidelity, post-class reflection architecture) and cultural substrate (collectivist versus individualist orientations). The pronounced geographic and methodological heterogeneity enables robust synthesis while foregrounding the necessity of culturally responsive design adaptations.

IV. DISCUSSION

Summary of Main Findings

The current systematic review compiled 83 empirical research studies which investigated how constructivist and SDL paradigms inform flipped classroom enactment in physical education. Three key trends were identified.

For one, concrete pedagogical enactments operationalized theoretical constructs rather than rhetoric as a gesture. Research that systematically aligned activities with Vygotskian ZPD learning and Zimmerman's (2000) SDL recursive model has produced significantly amplified effects (skill acquisition effect range $d = 0.71$ – 1.12) compared to atheoretical construct designs ($d = 0.35$ – 0.69). This is consistent with Hmelo-Silver et al.'s (2007) contention that pedagogical innovations succeed through theoretically coherent activity architectures supporting discrete cognitive processes.

Secondly, the tripartite phase structure opens up organic possibilities for constructivist and SDL integrations, but they are unevenly exploited. While pre-class video instruction was in fact almost universally adopted (90%), in-class peer practice was common (82%), yet post-class reflection, supposedly positioned to fully fulfill SDL regulatory cycles (Schön, 1983) became only 43% in implementation. Asymmetry of such a structure amounts to untapped potential: systemic post-class reflection associated with markedly increased SDL development ($d = 0.75$ – 0.94) as opposed to interventions that excluded reflective protocols ($d = 0.35$ – 0.58).

Third, cultural substrates moderate efficacy in paths that require culturally responsive calibration. A meta-analysis of 44 East and Southeast Asian implementations revealed collectivist learning dispositions require adaptive strategies. Successful implementations in Asia reconceptualized SDL as collective self-governance-groups working together to set goals and pursue retrospective analysis-rather than atomized individual autonomy. This adaptation is consistent with Markus and Kitayama's (1991) distinction between independent and interdependent self-construals, demonstrating SDL's metacognitive affordances are retained when integrated with collectivist axiological commitments.

Theoretical and Practical Implications

Theoretically, the results challenge a Western-centric understanding that characterizes SDL scholarship (Knowles, 1975) and show that autonomous learning is actually about being operationalized in structured collaborative design while, at the same time, not losing metacognitive affordances. This

carries SDL doctrine into collectivist epistemological areas, pointing to pedagogical contexts needing cultural translation instead of unmediated transfer.

Timeliness of intervention was a critical parameter: Developing competence skills in SDL depends on maintaining participation and sustained time of practice (≥ 10 weeks: $d = 0.78$ – 0.94), consistent with expertise theories which emphasise intensive deliberate practice (Ericsson, 2008). For practitioners, there are a few evidence-based imperatives crystallizing. First, pre-class video design must incorporate a combination of principles: expert modeling (Faul, 2007), self-assessment items, metacognitive prompts embedded into the video, and a time scale (8–15 min) ideally tailored to cognitive load requirements (Mayer, 2021). Second, in-class time intervals should be structured as deliberate practice with scaffolding, structured to gradually raise expectations. In explorations of systematic peer instruction - in which learners alternate instructor-performer roles according to explicit feedback criteria - effects were large ($d = 0.81$ – 0.95) in the context of studies like Østerlie and Mehus (2020). Third, SDL strategy instruction must be explicit, involving explicit goal articulation, self-monitoring routines, and reflective protocols delivered through modeling and guided rehearsal (Dignath & Büttner, 2008), with progressive responsibility transference as competence builds (Fisher & Frey, 2013). Facilitate collaborative goal-setting and collective reflective learning as well as individualised tasks for Asian implementations, and reestablish autonomy as shared accountability.

Constrained digital infrastructure necessitates early response through multimodal delivery: campus viewing stations, bandwidth-optimized variants, printing materials with QR code augmentation. Effective professional development is thus a must, focusing both on reconfiguration of pedagogical roles, video authorship, and interweaving of technology. Sun et al. (2017) showed that intensive preparation (20 hours) significantly increased implementation fidelity and outcomes compared to minimal training protocols.

Limitations and Future Directions

Several limitations need to be recognized. It is important to note that the heterogeneity reported in intervention formats and outcome instrumentation made synthesis challenging, precluding meta-analytic pooling-but narrative synthesis sufficiently addressed heterogeneity across intervention design/task (Popay et al., 2006). The majority analysed proximal outcomes, with only 30% employing follow-up for beyond 4 weeks, limiting conclusions on durability and SDL transfer between contexts. The constructivism-SDL analytic framework necessarily excluded different pertinent frameworks including cognitive load theory (Sweller et al., 2011) and self-determination theory (Ryan & Deci, 2000). Claims of cultural moderation are obtained indirectly from cross-study patterns, rather than from controlled cross-cultural experimental contrasts. Publishing bias can also exaggerate results: studies

who find themselves positively reported are better placed to win publication (see Rothstein et al., 2005).

Future research should focus on: (1) comparative effectiveness trials that isolate discrete design elements (e.g. peer instruction configuration or video formatting conventions) through component analysis frameworks (Levin et al., 2003); (2) longitudinal studies that track SDL transfer across learning ecologies and developmental trajectories using microgenetic approaches (Siegler, 2006); (3) mechanism studies in which mediation analysis is used to explicate causal paths between flipped architectural elements and outcomes (MacKinnon et al., 2007); (4) international collaborative research that implement standardized interventions across cultural contexts while systematically varying adaptation strategies to disentangle universal versus culturally contingent mechanisms; and (5) investigations of implementation science that examine factors that sustain scale-up and institutional embedding (Fixsen et al., 2005).

V. CONCLUSION

This systematic review shows how explicit implementation of constructivist and SDL paradigms substantially increases flipped classroom efficacy in physical education. An analysis of 83 studies discloses that theory-instituted implementations create moderate-to-large beneficial effects on student motivation ($g = 0.68$), skill acquisition ($g = 0.73$), knowledge retention ($g = 0.55$), and self-regulatory competence ($g = 0.71$). The combined framework provides evidence-grounded design heuristics for educators interested in optimizing flipped PE in diverse contexts. Importantly, results emphasise that effective implementation is more than simply "structural inversion" of content delivery—it demands deliberate operationalization of theoretical constructs through purposefully architected activities spanning all three flipped model phases.

Cultural substrates emerge as consequential mediators, in particular through Asian and Southeast Asian implementations where collectivist axiological commitments and pedagogical inheritances call for adaptive calibrations. Effective adaptations do not discard learner-centered pedagogy, but rather reconceptualise SDL in collaborative governance frameworks that pay homage to cultural dispositions while preserving theoretical fidelity. Given its continued international spread, sustained focus on culturally responsive pedagogical design and comprehensive professional development will likely be required in order to actualize the pedagogical potential generated by the proposed method, flipped classrooms.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the National Social Science Fund Project (Grant No. 23TY013).

DISCLOSURE

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The search strategy, screening framework, and data-extraction materials are available from the corresponding author on reasonable request.

ETHICS STATEMENT

Ethical approval and informed consent were not required because this study synthesised published literature only.

REFERENCES

1. Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
2. Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.
3. Casey, A., & Goodyear, V. A. (2015). Can cooperative learning achieve the four learning outcomes of physical education? A review of literature. *Quest*, 67(1), 56–72. <https://doi.org/10.1080/00336297.2014.984733>
4. Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research and Development*, 67(4), 793–824. <https://doi.org/10.1007/s11423-018-9633-7>
5. Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3(3), 231–264. <https://doi.org/10.1007/s11409-008-9029-x>
6. Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit.
7. Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, 15(11), 988–994. <https://doi.org/10.1111/j.1553-2712.2008.00227.x>
8. Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175–191.
9. Fisher, D., & Frey, N. (2013). *Better learning through structured teaching: A framework for the gradual release of responsibility* (2nd ed.). ASCD.
10. Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. University of South Florida, Louis de la Parte Florida Mental Health Institute.

11. Ghorbel, A., Trabelsi, O., Yaakoubi, M., Souissi, M. A., & Gharbi, A. (2025). Flipped classroom approach for gymnastics learning in physical education: A quasi-experimental study. *International Journal of Sports Science & Coaching*, 20(1), 300–312.
12. Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A meta-analysis. *BMC Medical Education*, 18(1), 38. <https://doi.org/10.1186/s12909-018-1144-z>
13. Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107. <https://doi.org/10.1080/00461520701263368>
14. Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/EFI-180221>
15. Hu, R., Gao, H., Ye, Y., Ni, Z., Jiang, N., & Jiang, X. (2018). Effectiveness of flipped classrooms in Chinese baccalaureate nursing education: A meta-analysis of randomized controlled trials. *International Journal of Nursing Studies*, 79, 94–103. <https://doi.org/10.1016/j.ijnurstu.2017.11.012>
16. Kirk, D. (2013). Educational value and models-based practice in physical education. *Educational Philosophy and Theory*, 45(9), 973–986. <https://doi.org/10.1080/00131857.2013.785352>
17. Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers*. Cambridge Adult Education.
18. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
19. Levin, J. R., O'Donnell, A. M., & Kratochwill, T. R. (2003). Educational/psychological intervention research. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 557–581). Wiley. <https://doi.org/10.1002/0471264385.wei0722>
20. MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614. <https://doi.org/10.1146/annurev.psych.58.110405.085542>
21. Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>
22. Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
23. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
24. Piaget, J. (1970). *Genetic epistemology*. Columbia University Press. <https://doi.org/10.7312/piag91272>
25. Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M. (2006). Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme. Lancaster University.
26. Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. <https://doi.org/10.1002/0470870168>
27. Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
28. Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
29. Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464–510). Wiley. <https://doi.org/10.1002/9780470147658.chpsy0211>
30. Østerlie, O., & Mehus, I. (2020). Effects of a flipped classroom intervention on sports education in physical education. *Journal of Teaching in Physical Education*, 39(1), 78–88. <https://doi.org/10.1123/jtpe.2018-0318>
31. Strelan, P., Osborn, A., & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, 30, 100314. <https://doi.org/10.1016/j.edurev.2020.100314>
32. Sun, J. C. Y., Wu, Y. T., & Lee, W. I. (2017). The effect of the flipped classroom approach to OpenCourseWare instruction on students' self-regulation. *British Journal of Educational Technology*, 48(3), 713–729. <https://doi.org/10.1111/bjet.12444>
33. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. <https://doi.org/10.1007/978-1-4419-8126-4>
34. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

35. Yoon, M., Hill, J., & Kim, D. (2021). Designing supports for promoting self-regulated learning in the flipped classroom. *Journal of Computing in Higher Education*, 33(2), 398–418. <https://doi.org/10.1007/s12528-021-09269-z>
36. Zainuddin, Z., & Halili, S. H. (2016). Flipped classroom research and trends from different fields of study. *International Review of Research in Open and Distributed Learning*, 17(3), 313–340. <https://doi.org/10.19173/irrodl.v17i3.2274>
37. Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–40). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>