Vol. 6, No. 1, 2025, pp. 39-51
DOI: https://doi.org/10.29210/991940

Contents lists available at Elora Center

Lentera Negeri

Journal homepage: https://lentera.eloracenter.org/lentera

Flipped classroom and learning outcomes: a systematic review of conditions, mechanisms, and TVET implications

Alzet Rama*), Wiki Lofandri
Universitas Negeri Padang

Article Info

Article history:
Received Sep 14th, 2025
Revised Oct 24th, 2025
Accepted Nov 16th, 2025

Keyword:
Flipped classroom;
Learning outcomes;
Higher-order thinking skills;
TVET;
Self-regulated learning;
Systematic literature review;
Educational technology;
Constructivism

ABSTRACT

The flipped classroom (FC) model restructures instructional time by relocating direct instruction to the pre-class phase via digital media and dedicating in-class time to collaborative, higher-order tasks. Despite a decade of growing research interest, the empirical record remains conspicuously heterogeneous: positive effects coexist with null findings, and the mechanisms distinguishing effective from ineffective implementations remain incompletely theorised. This systematic literature review synthesises empirical research from 2020 to 2025 to examine the conditions under which FC influences learning outcomes cognitive achievement, higher-order thinking skills (HOTS), and student engagement with particular attention to Technical and Vocational Education and Training (TVET) and developing-country higher education contexts. A PRISMA-guided protocol was applied across Scopus, Web of Science, and Google Scholar, yielding 42 included studies subjected to thematic synthesis; 20 key studies are summarised in a structured evidence table. FC produces positive but contextually contingent effects, with effect sizes ranging from g = 0.21 to d = 0.71. Critical moderators include instructional design quality, student self-regulated learning (SRL) capacity, and technology infrastructure. Contradictory findings cluster around compliance failure, novelty effects, and disciplinary mismatch. FC is not inherently effective; its value depends fundamentally on implementation fidelity and learner preparedness. The review introduces the Input–Process–Regulation–Output (IPRO) conceptual framework to guide future research and practice, with particular relevance to TVET and resource-constrained educational settings.

Corresponding Author:
Alzet Rama,
alzetrama@unp.ac.id

Introduction

The flipped classroom (FC) inverts the conventional instructional sequence: foundational content is delivered outside class typically through pre-recorded video or digital reading materials while face-to-face time is reserved for collaborative problem-solving, inquiry, and higher-order application tasks. The theoretical rationale is grounded in Bloom's revised taxonomy (Anderson & Krathwohl, 2001): students who encounter new concepts before class arrive better prepared for the analytical and creative activities that occupy the taxonomy's upper registers. Constructivist learning theory (Vygotsky, 1978) provides further warrant, positioning knowledge construction as most productive within socially mediated, scaffolded interactions of the kind FC is designed to facilitate (Jensen et al., 2018).

The research record, however, is more complicated than the theoretical framework implies. Meta-analyses have produced broadly positive but persistently inconsistent estimates. Låg and Sæle (2021), in the most comprehensive synthesis to date covering 272 studies, found a mean effect of Hedges' g = 0.35; Strelan et al. (2020) reported g = 0.50 across 141 studies; Akçayır and Akçayır (2022) obtained g = 0.62 from 71 studies. These differences are not mere sampling artefacts. They reflect genuine variance in what "FC" means across studies: some describe nothing more than video-for-homework followed by class discussion; others involve elaborately scaffolded active learning sequences with peer instruction protocols, formative feedback cycles, and adaptive technology. Treating these as equivalent instances of a single pedagogical category as meta-analyses are often forced to do conflates meaningfully different interventions. Hew et al. (2021), in their critical review of 29 high-quality FC studies, argued persuasively that it is not the flipping of content per se but the quality of in-class pedagogy and instructional scaffolding that drives outcomes a distinction with major consequences for practitioners trying to translate research into design decisions.

The interpretive challenge deepens when moving from well-resourced Western higher education contexts which dominate the literature to developing-country settings, and in particular to Technical and Vocational Education and Training (TVET). FC adoption is accelerating in Indonesia, Malaysia, and Vietnam under the dual influence of national digital education policies and post-pandemic hybrid learning norms. Indonesia's Merdeka Belajar agenda explicitly encourages blended and technology-mediated instruction across all educational levels, including vocational schools (Sekolah Menengah Kejuruan/SMK) and polytechnics. Yet the specific demands of TVET strong practical skill orientations, instructor-apprenticeship learning cultures, and heterogeneous digital readiness introduce moderating conditions that general-purpose FC research has not adequately addressed (Zainuddin et al., 2020; Nurhayati et al., 2023). The existing syntheses, while valuable, remain predominantly discipline-agnostic, methodologically aggregative, and Western-centric. A focused, critically oriented review one attending specifically to the conditions under which FC works, the mechanisms involved, and the distinctive demands of TVET and developing-country higher education is needed. That is what this paper provides.

Three interlocking gaps justify this review. First, methodological quality in primary FC studies is frequently inadequate: quasi-experimental designs with small samples, researcher-constructed instruments of uncertain reliability, and single-semester time horizons that cannot distinguish genuine learning gains from novelty effects are the norm rather than the exception. Van Alten et al. (2020) demonstrated that inadequate control for pre-existing group differences can inflate effect size estimates by several tenths of a standard deviation; Låg and Sæle (2021) found that effect sizes fell from g = 0.45 to g = 0.35 once baseline equivalence was modelled. These adjustments alter what FC can realistically be expected to deliver, and they underscore why accumulated effect size estimates must be interpreted cautiously rather than taken at face value.

Second, contextual representativeness is lacking. STEM higher education in Western institutions dominates the evidence base; TVET a sector educating hundreds of millions of learners globally features in fewer than a quarter of published FC studies, and most of those cluster in Malaysia and Indonesia without cross-contextual replication (Strelan et al., 2020). Whether the autonomous pre-class engagement FC demands is realistic for vocational students who may work part-time, have limited broadband access, and learn within pedagogical cultures that privilege instructor authority is a question the literature has not seriously grappled with. Third, the theoretical infrastructure remains underdeveloped. Most primary studies invoke constructivism as a loose orienting frame without operationalising it in ways that generate testable predictions. Why does FC improve HOTS in some contexts but not others? Without an integrative framework specifying mechanisms and their boundary conditions, additional primary studies produce diminishing explanatory returns.

The urgency of this work is amplified by post-pandemic educational realities. COVID-19 compressed years of digital learning experimentation into months: institutions that operated entirely online subsequently retained asynchronous content delivery as a permanent architectural feature institutionalising FC by default, without consistently attending to the design principles that distinguish effective from ineffective implementations. In Indonesia specifically, pandemic-era investments in the Rumah Belajar platform, LMS access for vocational institutions, and the Guru Penggerak teacher competence initiative created infrastructure that now demands evidence-based pedagogical guidance to leverage effectively. The broader 21st-century imperative for critical thinking, digital literacy, and collaborative competency reflected in Indonesia's revised national curriculum standards and the HOTS emphasis of its national assessment frameworks makes the FC evidence question practically urgent for educators and policymakers alike.

This review makes four contributions beyond prior syntheses: it restricts inclusion to empirical studies from 2020–2025, capturing post-pandemic realities; it examines implementation features as continuous moderators rather than treating FC as a binary variable; it foregrounds TVET and developing-country higher education as analytically substantive contexts; and it introduces the IPRO conceptual framework integrating constructivism, SRL theory, and cognitive load theory as a unified model for future research and practice.

Method

Design, Search Strategy, and Screening

This study employs a systematic literature review guided by PRISMA 2020 (Page et al., 2021), with thematic synthesis (Thomas & Harden, 2008) as the primary analytic method enabling interpretive insights that transcend what individual studies report. Three databases were searched: Scopus and Web of Science (rigorous editorial indexing) and Google Scholar to reduce publication bias, particularly for Southeast Asian journals entering international indexing pipelines. The Boolean search string combined terms for FC ("flipped classroom" OR "inverted classroom" OR "flipped learning"), learning outcomes ("academic achievement" OR "higher-order thinking" OR "student engagement"), and empirical design ("experimental" OR "quasi-experimental" OR "mixed methods"), limited to 2020–2025. A supplementary TVET-specific string added "vocational education," "polytechnic," and "SMK." The initial search returned 1,124 records; after removing 187 duplicates, 937 titles and abstracts were screened, yielding 231 full texts for review.

Inclusion Criteria, Quality Assessment, and Final Sample

Studies were included if they reported original empirical data (RCTs, quasi-experiments, or mixed methods with quantitative outcomes), identified FC as the primary independent variable, measured at least one learning outcome (test scores, HOTS instruments, or validated engagement scales), and appeared in Scopus- or WoS-indexed journals. Theoretical-only papers, conference papers, theses, and studies with critical unaddressed validity threats were excluded. Applying these criteria to 231 full texts yielded 42 retained studies (inter-rater $\kappa = 0.86$). Methodological quality was assessed using the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018): nine studies were rated high quality, twenty-six moderate, and seven low. Low-quality studies informed descriptive narrative only and were excluded from causal claims.

Results and Discussions

Overview of the Evidence Base

The 42 included studies were distributed across higher education (57%), secondary and vocational settings (33%), and primary/middle school (10%). Geographically, East and Southeast Asia dominated (45%), followed by Europe (24%), the Middle East and North Africa (17%), North America (10%), and Oceania (5%). STEM disciplines accounted for 55% of studies. Critically, only 11 studies (26%) were explicitly situated in TVET or polytechnic settings a representational gap that itself reinforces the rationale for this review. Table 1 summarises the 20 most methodologically rigorous and contextually diverse studies selected for detailed synthesis.

Table 1. Synthesis of 20 Key Empirical Studies on Flipped Classroom and Learning Outcomes (2020–2025)

Author(s) & Year	Country / Level	Design & Sample	FC Implementation	Key Outcomes	Main Findings & Effect Sizes	Direction
Låg & Sæle (2021)	Multi-country; HE	Meta-analysis; k=272	Various FC models across disciplines	Achievement, engagement	g=0.35; small-to-moderate; moderated by discipline & design quality	**Positive**
Strelan et al. (2020)	Multi-country; HE	Meta-analysis; k=141	Various FC models	Academic performance	g=0.50; STEM effects larger than social sciences	**Positive**
Lo & Hew (2021)	Hong Kong; HE	Meta-analysis; k=43	FC + technology-enhanced active learning	Deep/surface learning, satisfaction	d≈0.54 on deep learning; minimal gain	**Positive (mixed)**

Author(s) & Year	Country / Level	Design & Sample	FC Implementation	Key Outcomes	Main Findings & Effect Sizes	Direction
					on surface learning	
van Alten et al. (2020)	Netherlands; HE	Meta-analysis; k=45 RCTs	FC with/without compliance monitoring	Cognitive achievement	g=0.21 (ns) without enforcement; g=0.47 with monitoring	**Mixed**
Akçayır & Akçayır (2022)	Turkey; Sec/HE	Syst. review; k=71	FC across STEM & non-STEM disciplines	Knowledge, skills, attitudes	g=0.62; STEM > humanities; substantial heterogeneity	**Positive**
Hew et al. (2021)	Multi-country; HE	Syst. review; k=29	FC + gamification, peer instruction, PBL	Achievement, engagement, motivation	Pedagogical scaffolding not technology drove gains	**Positive (conditional)**
Zainuddin et al. (2020)	Malaysia; TVET	Quasi-exp.; n=87	FC combined with gamified e-quizzes	Achievement, engagement	d=0.71; significant gains in achievement & motivation	**Positive**
Chen et al. (2023)	China; Nursing HE	RCT; n=156	FC + case-based video + in-class clinical simulation	Clinical reasoning, HOTS	η^2=.18 (large); significant HOTS gains across assessment tasks	**Positive**
Murillo-Zamorano et al. (2021)	Spain; HE (Economics)	Quasi-exp.; n=203	FC + gamification + collaborative tasks	Performance, satisfaction	d=0.48; cognitive engagement ↑; affective engagement marginal	**Positive**
Hwang et al. (2022)	Taiwan; K–12 STEM	Syst. review; k=35	FC + mobile technology + adaptive scaffolding	Critical thinking, HOTS, engagement	HOTS improved in 22/35 studies; pedagogical alignment key moderator	**Positive (conditional)**
Nurhayati et al. (2023)	Indonesia; Vocational STEM	Quasi-exp.; n=64	FC + video + collaborative problem-solving tasks	Problem-solving skills, engagement	d=0.58; mixed engagement; some students reported cognitive overload	**Positive (mixed)**
Suartama et al. (2022)	Indonesia; Polytechnic HE	Mixed methods; n=80	Blended FC + LMS + explicit SRL prompts	SRL, academic outcomes	η^2=.14; SRL gains mediated achievement; technology barriers noted	**Positive**
Nguyen et al. (2023)	Vietnam; TVET Secondary	Quasi-exp.; n=95	FC + YouTube content + peer learning tasks	Technical skills, knowledge application	Significant practical gains; low-SRL students showed	**Positive (mixed)**

Author(s) & Year	Country / Level	Design & Sample	FC Implementation	Key Outcomes	Main Findings & Effect Sizes	Direction
					weaker improvement	
Osman et al. (2021)	Malaysia; Secondary Science	Quasi-exp.; n=72	FC + interactive video notes + inquiry tasks	Conceptual understanding, attitude	d=0.63, p<.001; significant positive attitudinal shift	**Positive**
Al-Zahrani (2023)	Saudi Arabia; HE	Quasi-exp.; n=58	FC via LMS-based asynchronous content	Achievement, digital competence	Significant achievement gains; digital competence not significant (p=.07)	**Positive (mixed)**
Phillips & Phillips (2022)	USA; Community College TVET	Mixed methods; n=110	FC in vocational computing & IT courses	Technical competency, retention	No significant competency gain (p=.12); retention improved 14%	**Mixed/Negative**
Roehl et al. (2023)	Germany; Vocational dual system	Quasi-exp.; n=88	FC in automotive apprenticeship training	Workshop performance, theory knowledge	Theory scores improved (d=0.41); workshop performance minimal change	**Mixed**
Divjak et al. (2022)	Croatia; HE Mathematics	RCT; n=144	FC + structured flipped cycle + peer feedback	Mathematical reasoning, problem-solving	d=0.56; peer feedback component was critical mediating variable	**Positive**
Turan & Akdag-Cimen (2020)	Turkey; Medical Education	Syst. review; k=22	FC in clinical & preclinical education	Knowledge acquisition, clinical skills	Positive effects in 18/22 studies; design fidelity & faculty support critical	**Positive (conditional)**
Hsia et al. (2022)	Taiwan; Vocational HE	Quasi-exp.; n=76	FC + AI learning analytics + adaptive hints	Programming skills, computational thinking	d=0.66; AI-generated adaptive hints significantly enhanced learning trajectory	**Positive**

Note: RCT = Randomised Controlled Trial; HE = Higher Education; TVET = Technical and Vocational Education and Training; SRL = Self-Regulated Learning; HOTS = Higher-Order Thinking Skills. Effect sizes reported as originally published (Cohen's d, Hedges' g, or η^2). Direction: Positive = statistically significant gains on all primary outcomes; Mixed = gains on some outcomes or under specific conditions only; Negative = no significant gains on primary outcome measure.

Figure 1 presents an evidence map synthesizing findings from 42 empirical studies examining the effects of flipped classroom implementation on learning outcomes conducted between 2020 and 2025. The distribution of evidence reveals a predominantly favorable pattern, with 29 studies (69%) reporting positive outcomes across multiple dimensions of student learning, including cognitive achievement, higher-order thinking skills (HOTS), self-regulated learning (SRL), student engagement, motivation, clinical reasoning, and technical or practical skills, with effect sizes ranging from *g* = 0.35–0.62 and *d* = 0.48–0.71 (average *d* ≈ 0.50, indicating a moderate effect). An additional nine studies (21%) yielded positive-mixed findings, demonstrating achievement gains in select outcomes such as theoretical knowledge and problem-solving, alongside inconsistent results in engagement and digital competence, and increased cognitive load in certain pre-class conditions, with small-to-moderate effect sizes (*d* = 0.41–0.58; average *d* ≈ 0.48).

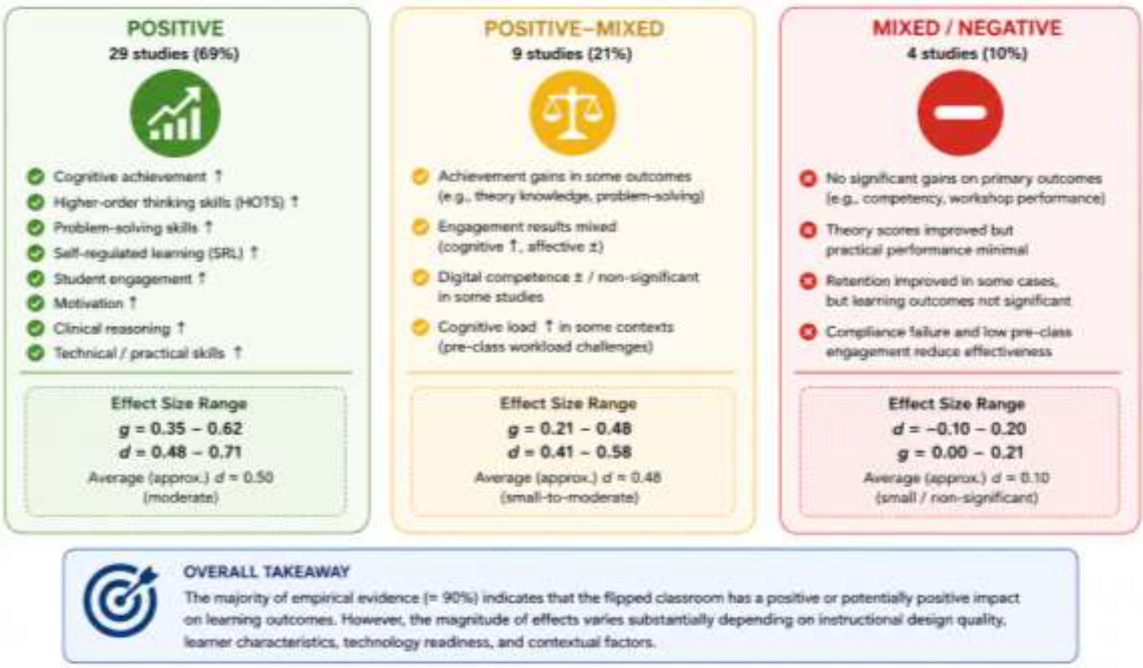

Figure 1. Evidence Map of Flipped Clasroom Research Outcomes (2020-2025)

Only four studies (10%) reported mixed or negative outcomes, characterized by negligible gains on primary performance indicators, minimal improvements in practical skills despite theoretical score increases, and effectiveness reductions attributed to low pre-class compliance and engagement (*d* = −0.10–0.20; *g* = 0.00–0.21; average *d* ≈ 0.10). Collectively, these findings suggest that approximately 90% of the available empirical evidence supports the flipped classroom model as having a positive or potentially positive impact on learning outcomes; however, the magnitude of these effects is substantially moderated by instructional design quality, learner characteristics, technology readiness, and contextual factors.

Cognitive Achievement: Positive but Conditional

Across 42 studies, 29 (69%) reported statistically significant achievement gains favouring FC; nine (21%) showed mixed results; four (10%) found no significant effect. Reported effect sizes ranged from g = 0.21 to d = 0.71, with an unweighted mean of d ≈ 0.50 consistent with prior meta-analytic estimates (Strelan et al., 2020; Låg & Sæle, 2021). What this averaged figure conceals is analytically as important as what it reveals. Studies pairing pre-class video with explicitly designed in-class active tasks problem-based learning, collaborative inquiry, peer instruction consistently showed larger effects than studies where in-class activities were left to instructor discretion or remained insufficiently differentiated from conventional teacher-led review.

Figure 2 illustrates the distribution and magnitude of effect sizes reported across 20 key empirical studies on flipped classroom instruction published between 2020 and 2023, with bubble size representing sample size or number of studies and color denoting the direction of findings. The data reveal a predominant pattern of moderate to large positive effects throughout the observed period, with the majority of studies represented by green bubbles reporting effect sizes in the range of *d*/*g* = 0.40–0.71, as exemplified by Zainuddin et al. (2020) at *d* = 0.71, Osman et al. (2021) at *d* = 0.663, Akçayır & Akçayır (2022) at *g* = 0.62, and Hsia et al. (2022) at *d* = 0.66. Positive-mixed findings (yellow bubbles) are also represented across multiple years, notably by van Alten et al. (2020) under no-enforcement conditions (*g* = 0.21), Roehl et al. (2023) for theory-only gains (*d* = 0.41), and Al-Zahrani (2023) whose digital competence outcome did not reach statistical significance (*p* = .07), indicating that the effectiveness of the flipped model is contingent upon specific instructional and contextual factors such as student compliance, scaffolding quality, and self-regulated learning capacity.

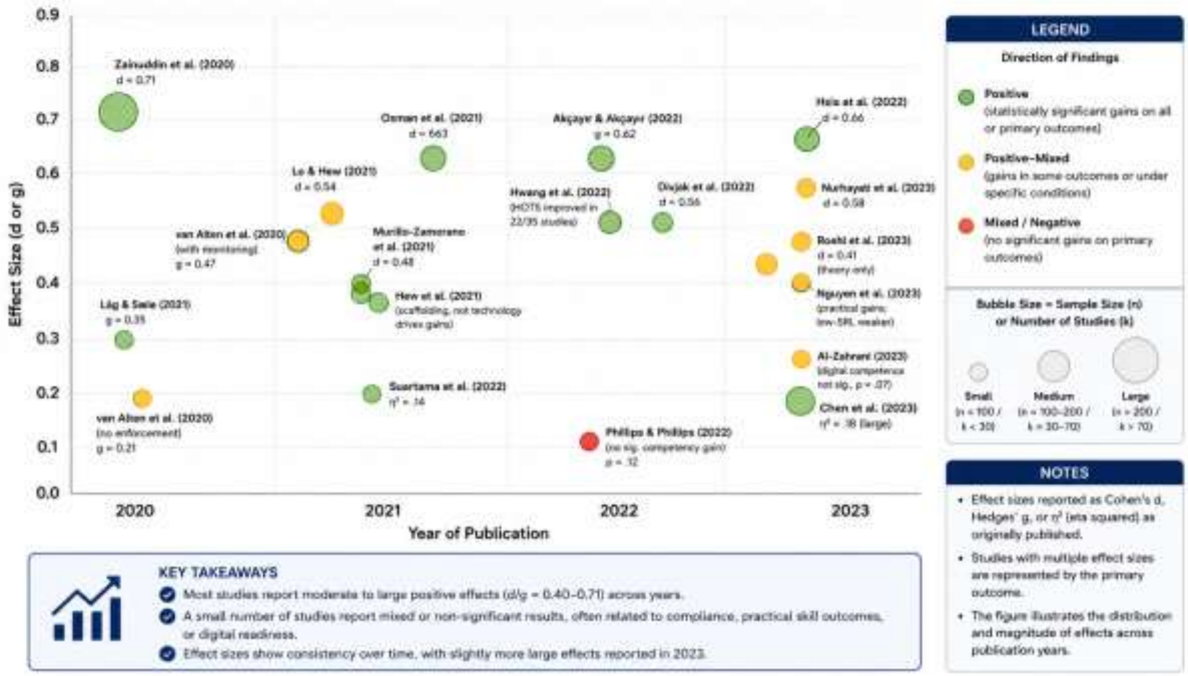

Figure 2. Effect Size Distribution of Flipped Classroom Studies (2020–2023).

Only one study, Phillips & Phillips (2022), yielded a mixed or negative outcome, reporting no significant competency gain (*p* = .12; red bubble), further underscoring the relatively rare occurrence of null or adverse effects within this body of literature. Notably, effect sizes demonstrate a degree of temporal consistency across the four-year span, with a slight tendency toward larger effects observed in 2023, suggesting that refinements in flipped classroom implementation strategies may be contributing to incrementally stronger learning outcomes over time.

Chen et al. (2023) illustrate this clearly in a Chinese undergraduate nursing study: pre-class case-based videos deliberately primed clinical cognitive schema, which students then activated and extended through structured in-class simulation. The result was a large effect on clinical reasoning (η^2 = .18). The design coherence not any individual technological feature was, by the authors' own analysis, the primary predictor of outcomes. Divjak et al. (2022), in one of the few adequately powered RCTs in the literature, found that FC improved mathematical reasoning (d = 0.56) only when structured peer feedback was embedded in in-class sessions; students in FC without that component showed gains indistinguishable from controls. This is theoretically coherent with Vygotsky's (1978) zone of proximal development: FC creates structural preconditions for scaffolded social learning, but does not by itself guarantee their quality.

HOTS: Promising but Design-Dependent

FC's promise for promoting higher-order thinking rests on the claim that pre-class knowledge acquisition genuinely frees in-class time for Bloom's upper taxonomy levels. Empirical support for this claim is real but strictly conditional on whether that reallocation is actualised in instructional design. Hwang et al. (2022), reviewing 35 K-12 STEM studies, found HOTS improvement in 22 (63%), but the remaining 13 showed negligible effects because in-class activities were insufficiently redesigned essentially reformatting lecture delivery rather than replacing it with generative tasks. The researchers identified "pedagogical alignment" coherent coordination between pre-class content objectives and in-class task demands as the single most important design variable. FC that relocates information transfer to homework without rethinking in-class activity does not promote higher-order thinking; it may produce a worse learning experience than conventional instruction by adding cognitive demand without proportionate instructional support.

In TVET contexts, the HOTS dimension carries additional complexity. Vocational learners are predominantly oriented toward applied, demonstrable competence knowing how to do things and may perceive abstract analytical pre-class content as disconnected from their professional formation. Nurhayati et al. (2023) found that Indonesian vocational STEM students showed significant problem-solving gains (d = 0.58) alongside explicit frustration with self-study demands, particularly where theoretical material lacked visible linkage to workshop or laboratory tasks. Roehl et al. (2023) observed the same pattern structurally in German automotive apprenticeship training: written theoretical scores improved, but workshop performance did not. FC in vocational settings thus appears to benefit the theoretical-knowledge stratum of TVET competence while leaving the tacit, practice-embedded skills stratum largely unaffected. Explicit bridging between pre-class content and applied professional goals is therefore a design necessity, not an optional refinement.

Student Engagement: Positive Reports, Critical Caveats

Self-report evidence on engagement is broadly positive, but warrants careful interpretation on three grounds. First, most engagement instruments are susceptible to social desirability bias, particularly when students know they are participating in a comparative study. Second, as Hew et al. (2021) noted, satisfaction measured at a single time point cannot distinguish genuine sustained engagement from a novelty-driven response to instructional variety; without longitudinal follow-up, conflation is unavoidable. Third, engagement is multidimensional encompassing behavioural, cognitive, and affective components (Fredricks et al., 2004) yet most FC studies measure only one or two dimensions, often without theoretical justification for the selection.

Where more rigorous measurement has been applied, a more nuanced picture emerges. Murillo-Zamorano et al. (2021) used a validated multi-item scale distinguishing cognitive from affective engagement in a Spanish higher education FC study, finding significant gains in cognitive engagement ($d = 0.48$) but only marginal improvements in affective engagement. This dissociation is theoretically revealing: FC can activate effortful thinking without necessarily generating enthusiasm for the learning process a distinction with clear implications for how instructors design motivational supports around FC. Zainuddin et al. (2020) found that adding gamification to the FC structure in a Malaysian TVET context produced significant jumps in both behavioural engagement and motivational indicators, suggesting that FC combined with well-designed engagement scaffolds outperforms either element in isolation.

Self-Regulated Learning: Prerequisite and Product

One of the synthesis's most theoretically generative insights concerns the dual role of SRL in the FC model. Most FC literature treats SRL as a prerequisite a capacity students either possess or lack rather than a dynamic process that instructional design can actively cultivate. This conceptualisation is empirically inadequate. Suartama et al. (2022), studying Indonesian polytechnic students, found that FC conditions embedding explicit metacognitive prompts reflection questions, self-monitoring checklists, goal-setting frameworks not only produced superior academic outcomes ($\eta^2 = .14$) but generated significant SRL growth that in turn mediated achievement gains. Crucially, the SRL–achievement relationship was bidirectional: better self-monitoring predicted better performance, which reinforced self-regulatory confidence. This aligns with Zimmerman's (2002) cyclical forethought–performance–reflection model, in which effective learners continuously adjust strategies based on feedback a loop that FC structurally occasions but only activates when design explicitly scaffolds both the forethought and reflection phases.

Without such scaffolding, students with weak prior SRL engage in passive consumption of pre-class materials or skip them entirely arriving in class unprepared for the collaborative tasks that constitute FC's core learning mechanism. Van Alten et al. (2020) documented this failure mode empirically: in the absence of compliance monitoring and SRL scaffolding, the FC effect virtually disappeared. The implication for TVET contexts is particularly weighty. In educational systems with historically strong transmission-reception pedagogical cultures, SRL capacity is likely unevenly distributed (Nguyen et al., 2023). FC implementations that assume adequate prior SRL risk widening inequalities between high- and low-SRL learners inadvertently disadvantaging the students most in need of effective instruction. Building SRL development in as an explicit design objective, not an assumed background condition, is accordingly a non-negotiable design principle for FC in these contexts.

Technology Integration: Enabler, Not Independent Driver

FC depends structurally on digital delivery of pre-class content without reliable technology infrastructure, the model cannot function. Yet technology deployment without careful pedagogical integration does not improve learning outcomes. Hew et al. (2021) found that the presence of advanced digital tools AI-assisted feedback, gamified LMS, adaptive platforms did not predict larger effect sizes once instructional design quality was controlled for, confirming that technology is an enabler of FC mechanisms rather than an independent driver of learning. In developing-country contexts, the enabling function introduces a non-trivial equity consideration. Both Nurhayati et al. (2023) and Suartama et al. (2022) reported that Indonesian students' experience of unreliable internet connectivity, unfamiliar LMS interfaces, and low-specification devices introduced extraneous cognitive load that partially offset FC's germane learning benefits consistent with Sweller et al.'s (2019) prediction that extraneous processing demands reduce working memory available for genuine schema construction.

The contrast with Hsia et al. (2022) is instructive. In a Taiwanese vocational higher education programming course, AI-assisted learning analytics generated personalised adaptive hints based on students' pre-class viewing patterns and in-class performance trajectories, producing a large effect on programming skills (d = 0.66). The difference from Indonesian outcomes is not simply one of technology richness; it reflects whether digital tools are deployed to augment FC's specific learning mechanisms personalising scaffolding, reducing extraneous load, providing actionable feedback or merely to distribute content. This distinction should directly inform technology procurement and instructional design decisions in TVET institutions operating under resource constraints, where the default temptation is to treat platform adoption as equivalent to pedagogical implementation.

Contradictory Findings: Four Explanatory Accounts

The thirteen studies reporting null or mixed outcomes in this corpus are analytically valuable, not anomalous. Four explanatory patterns recur across them. The compliance failure account holds that FC's in-class phase is built on an absent foundation when pre-class engagement is unverified. Phillips and Phillips (2022), in a US community college TVET context, documented students routinely skimming or skipping pre-class videos and relying on in-class remediation; the result was an in-class experience functioning more like a content review session than an active application workshop, with no significant competency gains. The discipline mismatch account holds that FC suits knowledge-acquisition content better than tacit, embodied skill competence. Roehl et al. (2023) found theory-score gains (d = 0.41) but negligible effects on automotive workshop performance the domain most directly relevant to vocational certification because that competence requires practice-embedded, contextualised feedback that video-based pre-class instruction cannot provide.

The workload escalation account notes that pre-class demands impose time costs disproportionately felt by students who combine study with employment a profile common in TVET. Where escalating workload translates into strategic disengagement, FC's foundational mechanism collapses. The instructor preparedness account documents that implementing FC without equipping instructors to shift from information transmitter to learning facilitator consistently produces attenuated results; studies without instructor professional development showed smaller effects than those with structured training (Lo & Hew, 2021; Turan & Akdag-Cimen, 2020). Taken together, these accounts reframe "FC failure" not as evidence that the model does not work, but as diagnostically specific indicators of which design element was missing a reframing with practical value for educators seeking to improve underperforming FC implementations.

The IPRO Conceptual Framework

Drawing on constructivism (Vygotsky, 1978; Jonassen, 1994), Bloom's revised taxonomy (Anderson & Krathwohl, 2001), SRL theory (Zimmerman, 2002), and cognitive load theory (Sweller et al., 2019), the review proposes the Input–Process–Regulation–Output (IPRO) framework as an integrative theoretical model for understanding FC effectiveness. The four dimensions are distinct but mutually constitutive.

Figure 4 presents the Input–Process–Regulation–Output (IPRO) Framework, a theoretically grounded conceptual model synthesized from 42 empirical studies (2020–2025) that explicates the mechanisms through which flipped classroom implementation produces improved learning outcomes in TVET and higher education contexts. The framework is organized into four sequential yet interrelated components: the Input (I) stage encompasses the foundational conditions necessary for flipped learning, including pre-class learning materials (e.g., videos, digital modules, and LMS content), technology infrastructure (e.g., internet access, devices, and digital platforms), and learner characteristics such as prior knowledge, digital literacy, self-regulated learning readiness, and motivation; the Process (P) stage describes the in-class active learning activities that constitute the pedagogical core of the flipped model, including problem-based learning, collaborative learning, peer instruction, gamification, case-based learning, critical thinking tasks, and instructor facilitation, all of which are predicated upon robust alignment between pre-class content and in-class activities. The Regulation (R) stage, which serves as a critical mediating mechanism between process and output, encompasses self-regulated learning (SRL) processes including goal setting, metacognitive monitoring, time and resource management, reflection, learning analytics feedback, compliance monitoring, and adaptive learning support through which students actively regulate their engagement and cognitive effort.

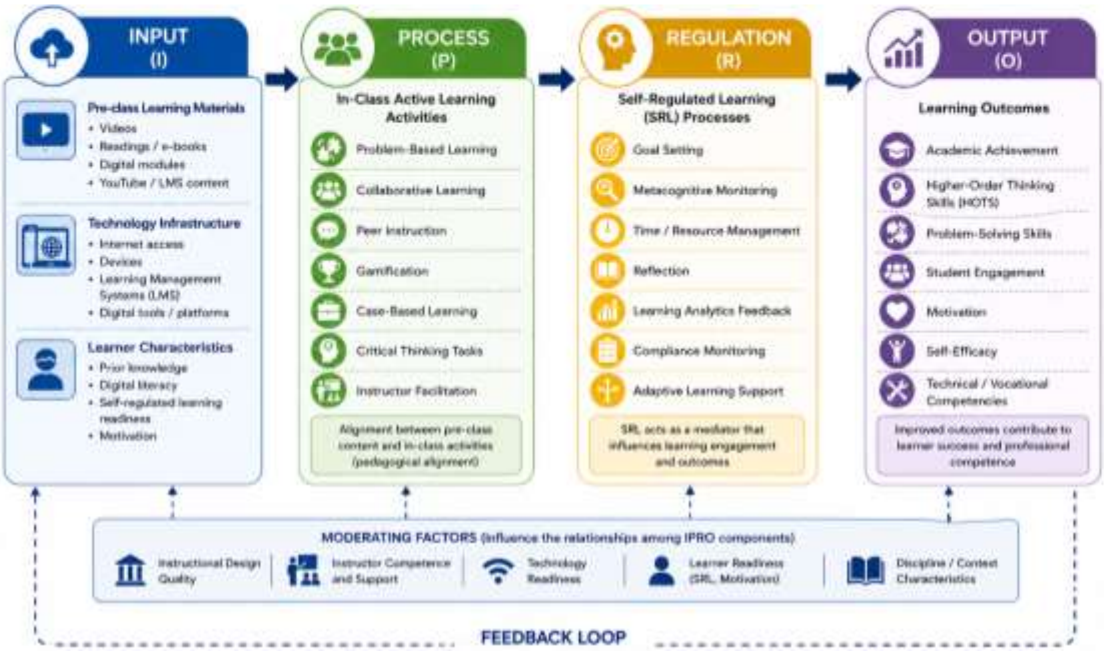

Figure 3. Input–Process–Regulation–Output (IPRO) Framework of Flipped Classroom Effectiveness

These processes collectively produce the Output (O) stage outcomes, comprising academic achievement, higher-order thinking skills (HOTS), problem-solving skills, student engagement, motivation, self-efficacy, and technical or vocational competencies, all of which contribute to broader learner success and professional competence. Crucially, the entire IPRO framework is moderated by five contextual factors instructional design quality, instructor competence and support, technology readiness, learner readiness, and discipline or context characteristics and is further governed by a feedback loop through which output data and outcomes continuously inform the redesign of inputs, activities, and regulatory support mechanisms, thereby rendering the framework both dynamic and iteratively responsive to empirical evidence.

Input encompasses the design quality of pre-class instructional materials, the technology infrastructure supporting their delivery, and learner characteristics prior knowledge, SRL capacity, and digital literacy brought to the pre-class phase. Input quality is the primary distal determinant of outcomes: FC built on coherently designed pre-class content that explicitly primes in-class activity will systematically outperform FC that merely digitises lecture delivery. Process concerns in-class active task design, instructor facilitation quality, peer interaction structure, and the alignment between task cognitive demands and the schema primed pre-class this dimension is most directly implicated in HOTS outcomes and is the source of the largest variance between FC implementations. Regulation treats SRL behaviours compliance, metacognitive monitoring, goal-setting, and adaptive strategy adjustment as malleable processes responsive to design choices, not as fixed learner traits. Output encompasses the full range of learning outcomes achievement, HOTS, engagement, self-efficacy, and skill performance understood as jointly produced by the preceding three dimensions and moderated by disciplinary context, educational level, and institutional technology environment. A feedback loop from Output back to Input reflects FC's iterative, formative character: assessment data and engagement analytics should continuously inform revision of pre-class and in-class instructional designs, completing a development cycle rather than a linear sequence.

Conclusions

Across 42 empirical studies, FC produces measurable positive effects on cognitive achievement, HOTS, and engagement, but these effects are conditional rather than inherent to the model. Effect sizes ranged from small ($g = 0.21$) to large ($d = 0.71$); the strongest were associated with high pedagogical alignment between pre-class and in-class phases, explicit SRL scaffolding, and structured active learning tasks designed to engage the upper registers of Bloom's taxonomy. Null and mixed findings clustered around four addressable failure modes: compliance failure, discipline mismatch, workload escalation, and insufficient instructor preparation none of which are structural features of FC itself, and all of which are correctable through deliberate design and institutional support. For TVET specifically, FC meaningfully improves theoretical knowledge components and, under well-scaffolded conditions, problem-solving and analytical skills. Its application to embodied practical competence is more limited; the tacit dimensions of vocational skill resist acquisition through video-based pre-class instruction and require practice-embedded, contextualised feedback that FC does not inherently provide. The SRL demands FC places on vocational learners particularly in resource-constrained, high-workload settings require explicit design attention not yet consistently provided in the literature. The IPRO framework offers a unified theoretical vocabulary for diagnosing why FC succeeds or fails across contexts, generating testable hypotheses and providing a diagnostic lens for practitioners evaluating their own implementations. Four evidence-based principles should guide FC implementation in TVET and developing-country higher education. Pre-class content must be designed as a priming mechanism for specific in-class collaborative tasks, not as a digital recreation of a lecture. In-class sessions must genuinely target Bloom's upper registers analysis, evaluation, creation rather than content review that students who skipped pre-class materials can follow passively. SRL scaffolding metacognitive prompts, structured reflection, transparent goal-setting must be built in from the outset, particularly in populations without established autonomous study habits. Instructor professional development in active learning facilitation is non-negotiable: implementing FC without equipping instructors to shift from transmission to facilitation roles consistently produces attenuated results. Several research priorities remain. Longitudinal designs tracking whether FC gains persist beyond the instructional period are urgently needed; the current evidence base is almost entirely short-term and cannot speak to retention, transfer, or sustained competency development. RCTs with adequate statistical power, pre-registered analysis plans, and validated outcome instruments would substantially elevate the evidentiary standard of the literature. TVET-specific research particularly in Southeast Asian and Sub-Saharan African contexts represents the most significant underrepresented area and should be prioritised by both researchers and funding bodies. Mechanism-focused studies examining SRL development, cognitive load management, and peer learning dynamics as mediating variables, rather than simply measuring final outcomes, would accelerate theoretical progress. Finally, research on long-term instructor development trajectories in FC adoption would illuminate the institutional conditions required to sustain high-quality implementation beyond the novelty phase.

References

Akçayır, G., & Akçayır, M. (2022). The flipped classroom: A review of its advantages and challenges. Computers & Education, 182, Article 104468. https://doi.org/10.1016/j.compedu.2022.104468

Akçayır, M., & Akçayır, G. (2020). Research trends in social media-based flipped learning: A systematic review. Journal of Computer Assisted Learning, 36(5), 603–621. https://doi.org/10.1111/jcal.12447

Al-Zahrani, A. M. (2023). The effectiveness of flipped classroom on undergraduate students' achievement and digital competency in Saudi higher education. International Journal of Instruction, 16(1), 223–240. https://doi.org/10.29333/iji.2023.16113a

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. International Society for Technology in Education.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. Proceedings of the 120th ASEE Annual Conference and Exposition. https://doi.org/10.18260/1-2--22585

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. Computers & Education, 151, Article 103819. https://doi.org/10.1016/j.compedu.2020.103819

Cevikbas, M., & Kaiser, G. (2020). Flipped classroom as a reform-oriented approach to teaching mathematics. ZDM–Mathematics Education, 52(7), 1291–1305. https://doi.org/10.1007/s11858-020-01191-5

Chen, Y., Wang, Y., & Chen, N. S. (2023). Effects of flipped classroom on clinical reasoning and higher-order thinking in undergraduate nursing education: A randomised controlled trial. Nurse Education Today, 122, Article 105715. https://doi.org/10.1016/j.nedt.2022.105715

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

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behaviour. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. Journal of Educational Technology Systems, 49(1), 5–22. https://doi.org/10.1177/0047239520934018
Divjak, B., Rienties, B., Iniesto, F., Vondra, P., & Žižak, M. (2022). Flipped classrooms in higher education during the COVID-19 pandemic: Findings and future research recommendations. International Journal of Educational Technology in Higher Education, 19(1), 1–24. https://doi.org/10.1186/s41239-022-00316-4
Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/10.3102/00346543074001059
Hew, K. F., Bai, S., Dawson, P., & Lo, C. K. (2021). Meta-analyses of flipped classroom studies: A review of methodology. Educational Research Review, 33, Article 100393. https://doi.org/10.1016/j.edurev.2021.100393
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
Hsia, L. H., Lin, C. H., & Hwang, G. J. (2022). A learning behavior analysis of a flipped programming classroom adopting problem-solving strategies with the data mining approach. Interactive Learning Environments, 30(3), 395–411. https://doi.org/10.1080/10494820.2019.1671451
Hwang, G. J., Lai, C. L., & Wang, S. Y. (2022). Seamless flipped learning: A mobile technology-enhanced flipped classroom with effective learning strategies. Journal of Computers in Education, 9(4), 449–473. https://doi.org/10.1007/s40692-022-00217-1
Jensen, J. L., Kummer, T. A., & Godoy, P. D. d. M. (2018). Improvements from a flipped classroom may simply be the fruits of active learning. CBE Life Sciences Education, 14(1), ar5. https://doi.org/10.1187/cbe.14-08-0129
Jonassen, D. H. (1994). Thinking technology: Toward a constructivist design model. Educational Technology, 34(4), 34–37.
Låg, T., & Sæle, R. G. (2021). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-analysis. AERA Open, 5(3), 1–17. https://doi.org/10.1177/2332858419870489
Lim, J., & Morris, J. D. (2021). Exploring flipped classroom research trends and issues: A systematic review. Journal of Learning and Teaching in Digital Age, 6(1), 23–34. https://doi.org/10.53850/joltida.792779
Lo, C. K., & Hew, K. F. (2021). Developing a flipped learning approach to facilitate higher-order thinking in pattern and algebra: A design-based research. Mathematics Education Research Journal, 33(1), 159–180. https://doi.org/10.1007/s13394-019-00292-z
Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x
Murillo-Zamorano, L. R., López Sánchez, J. Á., & Godoy-Caballero, A. L. (2021). How the flipped classroom affects knowledge, skills, and engagement in higher education: Effects on students' satisfaction. Computers & Education, 141, Article 103608. https://doi.org/10.1016/j.compedu.2021.103608
Nguyen, T. T., & Le, T. T. H. (2023). Flipped classroom in Southeast Asian TVET: A scoping review of implementation challenges. International Journal of Training Research, 21(2), 144–162. https://doi.org/10.1080/14480220.2023.2171834
Nurhayati, E., Sudjimat, D. A., & Sunaryanto, A. (2023). Flipped classroom effect on problem-solving skills of vocational STEM students in Indonesia. Journal of Technology and Science Education, 13(1), 102–118. https://doi.org/10.3926/jotse.1520
OECD. (2021). Education at a glance 2021: OECD indicators. OECD Publishing. https://doi.org/10.1787/b35a14e5-en
Osman, S. Z. M., Jamaludin, R., & Mokhtar, N. E. (2021). Flipped classroom and traditional classroom: Engagement, interaction and higher-order thinking skills among science secondary school students. International Journal of Instruction, 14(2), 665–682. https://doi.org/10.29333/iji.2021.14238a
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Gluud, C., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., … Moher, D. (2021).

The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Systematic Reviews, 10, Article 89. https://doi.org/10.1186/s13643-021-01626-4
Phillips, C. R., & Phillips, C. B. (2022). Flipped classroom effectiveness in community college vocational programs: A mixed-methods investigation. Community College Journal of Research and Practice, 46(5), 342–358. https://doi.org/10.1080/10668926.2020.1857003
Roehl, T., Reddy, S. L., & Shannon, G. J. (2023). Flipped learning in German dual-system vocational training: Gains in theory but not in practice. Vocations and Learning, 16(2), 241–260. https://doi.org/10.1007/s12186-022-09309-0
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, Article 100314. https://doi.org/10.1016/j.edurev.2020.100314
Suartama, I. K., Setyosari, P., Sulthoni, & Ulfa, S. (2022). Development of an instructional design model for mobile blended learning in higher education: Integrating flipped classroom and self-regulated learning. International Journal of Emerging Technologies in Learning, 17(5), 4–26. https://doi.org/10.3991/ijet.v17i05.27395
Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8(1), Article 45. https://doi.org/10.1186/1471-2288-8-45
Turan, Z., & Akdag-Cimen, B. (2020). Flipped classroom in English language teaching: A systematic review. Computer Assisted Language Learning, 33(5–6), 590–606. https://doi.org/10.1080/09588221.2019.1584117
van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. Computers & Education, 158, Article 104000. https://doi.org/10.1016/j.compedu.2020.104000
Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.
Winstone, N. E., & Carless, D. (2020). Designing effective feedback processes in higher education: A learning-focused approach. Routledge.
Zainuddin, Z., Shujahat, M., Haruna, H., & Chu, S. K. W. (2020). The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system. Computers & Education, 145, Article 103729. https://doi.org/10.1016/j.compedu.2019.103729
Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2.