Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Evidence-Based Quality Indicators For Flipped Classroom Teacher Development in Mathematics: A Meta-Analysis Informing Higher Education Practice

Luisa Morales-Maure[1,7], Evelyn García Vazquez[2], Orlando García-Marimón[3,7]*, Keila Chacón-Rivadeneira[4], Berenice Alfaro-Ponce[5], Y Miguel Ángel Esbrí Montoliu[6, 7]

1. Universidad de Panamá; luisa.morales@up.ac.pa, https://orcid.org/0000-0003-3905-9002
2. Universidad de Panamá; evelyn.garcia-v@up.ac.pa, https://orcid.org/0000-0002-0168-5350
3. Universidad de Panamá; orlando.egarcia@up.ac.pa, https://orcid.org/0000-0002-0334-6133
4. ITSE[1]; kchacon@itse.ac.pa, https://orcid.org/0009-0004-5967-4613.
5. Tecnológico de Monterrey, berenice.alfaro@tec.mx, https://orcid.org/0000-0002-7046-1019
6. Universidad de Panamá; miguel.esbri@up.ac.pa, https://orcid.org/0000-0002-1730-7661
7. National Research System, SENACYT, Panama City, Panama

*Corresponding Author: E-mail: orlando.egarcia@up.ac.pa

Abstract
This study develops evidence-based quality indicators for flipped classroom teacher development in mathematics by integrating a systematic review, quantitative meta-analysis, and thematic synthesis of national and international research. Following PRISMA 2020, 580 records were identified, 478 unique records were screened, and 70 studies were retained for the qualitative synthesis. A random-effects meta-analysis of 19 studies with recoverable uncertainty estimates yielded a statistically significant moderate effect on mathematics learning ($d = 0.522$, $SE = 0.147$, 95% CI [0.233, 0.811], $p = .0004$). Substantial heterogeneity was observed, $Q(18) = 208.089$, $p < .001$, $I^2 = 91.35\%$, and $\tau^2 = 0.369$, indicating that effectiveness varied markedly across educational levels, mathematical domains, instructional designs, and implementation contexts. Convergent evidence from 46 primary- and secondary-level effects reported by Yakar showed a comparable pooled effect ($d = 0.510$, 95% CI [0.340, 0.690], $I^2 = 87.91\%$). The thematic synthesis identified seven interdependent quality indicators: intentional pre-class design, mathematically rich in-class activity, formative assessment and responsive feedback, integrated mathematical-pedagogical-technological knowledge, explicit support for self-regulated learning, equitable digital access, and analytics-informed instructional adaptation. The findings show that flipped learning is not effective through content relocation alone; its educational value depends on coherence among mathematical tasks, teacher expertise, learner preparation, assessment, technology, equity, and reflective use of learning data. The proposed framework offers a transferable basis for designing, monitoring, and evaluating mathematics teacher-development programmes and higher education practice.

Keywords - Flipped classroom, mathematics education, teacher professional development, meta-analysis, quality indicators, learning analytics, self-regulated learning.

1. Introduction

Education systems are increasingly expected to improve learning outcomes while narrowing long-standing achievement gaps, a goal that becomes more complex in settings characterised by geographic dispersion, unequal access to educational resources, and interruptions that weaken instructional continuity. In the Panamanian context, these structural constraints strengthen the case for instructional approaches that are evidence-based, scalable across diverse school ecologies, and robust under real implementation conditions, rather than relying on innovation claims that lack conceptual and evaluative

[1] Specialized Higher Technical Institute of Panama

rigor (Darling-Hammond et al., 2017; Moon et al., 2026). Within this landscape, teacher education and ongoing professional development (PD) remain pivotal levers because instructional improvement depends on teachers' capacity to integrate disciplinary knowledge, pedagogical judgement, and insight into students' typical learning difficulties. Foundational perspectives on teacher knowledge emphasise that effective teaching requires specialized content knowledge for teaching and the professional ability to transform subject matter into learnable forms (Shulman, 1986; Ball et al., 2008). Accordingly, large-scale syntheses of PD indicate that impact is strongest when learning is sustained, situated in practice, supported through feedback, and embedded in collaborative structures rather than delivered as isolated workshops (Darling-Hammond et al., 2017).
Among the pedagogical innovations increasingly adopted in teacher education, the flipped classroom (FC) has gained prominence for reorganising learning time: students engage with core content before class through digital resources, and face-to-face time is repurposed for higher-order activity such as problem solving, discussion, formative assessment, and feedback. Meta-analytic findings indicate that flipped classroom (FC) approaches are associated with improvements in achievement and learner attitudes; however, reported effects are heterogeneous and depend on study design, educational level, and implementation characteristics, underscoring the need for context-sensitive interpretation rather than generalised claims of effectiveness (Yang & Chen, 2026). This heterogeneity is particularly consequential when FC is framed as a lever for mathematics teacher professional development, given that students' opportunities to reason, justify, and engage with cognitively demanding tasks are mediated by teachers' orchestration of classroom discourse, representations, and feedback. In parallel, professional development research cautions against accelerating "innovation" narratives without commensurate conceptual precision and evaluative robustness; active-learning reforms should therefore be grounded in explicit theory, supported by rigorous evidence, and examined through the perspectives of those who enact them, so that novelty is not conflated with quality (Moon et al., 2026). For FC-based teacher development, this implies that effectiveness should be judged not only by outcome gains but by transparent, evidence-based quality indicators that specify what "good" implementation entails in mathematics teaching contexts.

Learning analytics offers a complementary pathway for specifying and validating such quality indicators because it operationalises learning and teaching processes through traceable data from learners and learning environments. As a field, learning analytics integrates measurement, computational techniques, and pedagogical interpretation to understand learning, identify risk patterns, and support data-informed instructional decisions (Ferguson, 2012; Siemens & Baker, 2012). In flipped-learning contexts, scoping and empirical work shows that analytics can help teachers examine students' preparation, engagement trajectories, and in-class participation, thereby supporting adaptation of instructional design (Algayres & Triantafyllou, 2020; Fatmiyati et al., 2026). However, evidence also indicates that dashboards and reports only become actionable when teachers can interpret data meaningfully and align it with pedagogical goals, a capacity that is uneven and itself a PD target (van Leeuwen, 2019; Fatmiyati et al., 2026). These findings motivate the need to articulate FC quality indicators that are both empirically grounded and usable by teachers and teacher educators—indicators that connect observable behaviours and analytics traces to the professional knowledge base required for effective mathematics teaching.

This article addresses these challenges by developing evidence-based quality indicators for flipped classroom teacher development in mathematics through a systematic review and meta-analysis, with explicit implications for higher education practice. The study synthesises national and international evidence on flipped classroom effectiveness and identifies which implementation features and evaluative criteria show the most consistent associations with improved outcomes in mathematics learning contexts. In this way, it responds to calls to strengthen the conceptual and methodological foundations of innovation claims in professional development research (Moon et al., 2026) and is informed by evidence that structured improvement approaches in teacher education, including lesson study and didactic suitability criteria, can support teachers' reflective judgement and guide pedagogical innovation in higher education mathematics programmes (Morales-Maure et al., 2025a). The study is guided by two research questions: (RQ1) What evidence exists, nationally and internationally, regarding the effects of flipped classroom approaches on mathematics learning in primary and secondary education when implemented through teacher professional development? (RQ2) What evidence-based quality indicators for flipped-classroom teacher development in mathematics emerge from this literature, and how can these indicators inform higher education practices for designing, monitoring, and evaluating teacher learning?

2. Theoretical Framework

This theoretical framework integrates the main conceptual perspectives required to explain the effectiveness and variability of flipped classroom interventions in mathematics education and teacher professional development. Rather than treating flipped learning as a technological arrangement, the framework approaches it as a complex pedagogical model shaped by the interaction among mathematical content, instructional design, teacher knowledge, learner self-regulation, classroom participation, digital technologies, and evidence-informed decision-making. It draws on teacher professional knowledge, constructivism, active learning, self-regulated learning, TPACK, learning analytics, and didactical suitability to establish the analytical dimensions through which the quality of flipped mathematics instruction can be examined. This integrated perspective is particularly relevant for interpreting heterogeneous outcomes across studies and for identifying the conditions under which flipped learning contributes to meaningful, equitable, and sustainable improvements in mathematics teaching and learning.

2.1. Flipped Classroom as a Pedagogical Reorganisation of Mathematics Teaching

The flipped classroom should not be understood merely as the transfer of lectures from the classroom to digital environments. Its defining feature is the deliberate redistribution of instructional activity across pre-class and in-class phases so that classroom time can be devoted to higher-order mathematical reasoning, formative assessment, collaborative problem solving, and teacher-mediated discussion. In this sense, the flipped classroom represents a pedagogical reorganisation rather than a technological format. Recorded explanations, digital readings, interactive simulations, and online quizzes may facilitate initial exposure to mathematical content, but they do not constitute flipped learning unless they are coherently connected to meaningful classroom activity.

Research in mathematics education indicates that the effectiveness of this model depends on the quality of the relationship between both phases. Studies conducted in secondary and higher education have associated flipped instruction with gains in conceptual understanding, achievement, engagement, and attitudes toward mathematics, although effects vary considerably across contexts and implementation designs (Bhagat et al., 2016; Surjono & Mahmudi, 2023; Karjanto & Acelajado, 2022; Love et al., 2014; Wei et al., 2020). Meta-analytic evidence similarly suggests an overall positive effect on mathematics performance, but also reveals substantial heterogeneity, indicating that the instructional label alone does not explain effectiveness (Güler et al., 2023; Haziki et al., 2025; Yakar, 2021).

This variability is particularly important because apparently similar flipped interventions may differ in the duration and accessibility of pre-class materials, the degree of accountability for preparation, the mathematical demand of classroom tasks, the amount of feedback provided, and the teacher's capacity to respond to students' reasoning. Cevikbas and Kaiser (2020) therefore conceptualise flipped mathematics instruction as a reform-oriented approach whose value lies in creating conditions for student-centred learning, mathematical communication, differentiation, and active participation. From this perspective, inversion is not an end in itself; it is an organisational mechanism that can expand opportunities for mathematically productive activity when supported by appropriate pedagogical design.

The findings of Atta and Bonyah (2023) illustrate this principle in the preparation of future teachers. Their work on plane geometry suggests that flipped instruction can support both mathematical learning and professional preparation when pre-service teachers engage with representations and explanations before class and subsequently apply them through guided tasks. Likewise, Sacristán San Cristóbal et al. (2017) showed the relevance of flipped mathematics didactics in online teacher education, where the model enabled prospective early-childhood teachers to connect theoretical preparation with the analysis and design of instructional practice. These findings support the view that flipped learning in teacher education has a dual function: it develops participants' mathematical and didactical knowledge while simultaneously modelling an instructional approach that they may later adapt in their own classrooms.

2.2. Teacher Professional Development and Professional Knowledge for Flipped Mathematics Instruction
Teacher professional development is most effective when it is sustained, content-focused, connected to practice, collaborative, and organised around evidence of student thinking. For mathematics educators, professional learning must move beyond general digital skills and address the specialised knowledge required to select mathematical tasks, anticipate strategies, interpret misconceptions, sequence representations, and respond productively to incomplete or unexpected reasoning.

Shulman's concept of pedagogical content knowledge locates professional expertise in the transformation of disciplinary knowledge into representations, analogies, explanations, questions, and tasks that make the content accessible to particular learners (Shulman, 1986, 1987). Ball et al. (2008) extended this logic through the construct of mathematical knowledge for teaching, which includes specialised forms of mathematical understanding used to evaluate non-standard solutions, interpret students' errors, select examples, identify mathematically significant differences between strategies, and determine whether an explanation is complete and valid.

These domains are central to flipped mathematics education. The design of pre-class materials requires teachers to determine which ideas can reasonably be introduced independently, which representations may reduce or increase conceptual difficulty, and what forms of guidance students need to interpret mathematical notation, diagrams, procedures, or proofs. The in-class phase requires an even more complex form of professional judgement because teachers must use evidence from students' preparation to adapt explanations, organise groups, select solution methods for public discussion, and connect informal reasoning with formal mathematical structures.

The role of the teacher therefore becomes more demanding rather than less important. In well-designed flipped instruction, the teacher functions as a curriculum designer, diagnostician, facilitator of discourse, provider of feedback, and interpreter of digital and classroom evidence. Toivola et al. (2023) found that mathematics teachers' rationales for flipped learning were related to personalisation, student responsibility, participation, and the reorganisation of classroom interaction. However, teachers' accounts also show that these purposes require continuous pedagogical monitoring. The availability of videos or digital resources does not automatically generate student autonomy; teachers must design structures that make preparation meaningful and connect it explicitly to subsequent mathematical activity. Webel et al. (2018) similarly demonstrated that implementing flipped instruction in elementary mathematics involves complex adaptations. Their case study showed that the success of the approach depended on how the teacher interpreted student needs, managed access to instructional resources, and used classroom time for differentiated support. This reinforces the importance of professional development that helps teachers examine not only which technologies to use but also why, when, and for whom a particular instructional decision is appropriate. For this reason, professional development for flipped mathematics instruction should include cycles of design, implementation, evidence collection, collaborative analysis, and redesign. These cycles are consistent with reflective practice and Lesson Study because they position the classroom as a site of systematic professional inquiry. Teachers can jointly examine the mathematical and cognitive demands of pre-class resources, anticipate student responses, analyse learning traces and classroom work, and revise the lesson in light of observed evidence. Under this model, professional learning is directly connected to practice rather than separated from it.

2.3. Constructivism, Active Learning, and Mathematical Meaning-Making

Constructivist perspectives provide a theoretical explanation for why the redistribution of instructional time may improve learning. Mathematical knowledge is not acquired through passive exposure alone; it is constructed as learners coordinate prior knowledge, representations, procedures, language, examples, and feedback. Initial exposure to content may occur asynchronously, but conceptual development depends on opportunities to test conjectures, compare strategies, justify relationships, resolve inconsistencies, and reorganise understanding.

The classroom phase is consequently the mechanism through which inversion becomes mathematical learning. Carefully sequenced problems, productive struggle, peer explanation, teacher questioning, multiple representations, and formative assessment enable learners to transform preliminary information into connected mathematical knowledge. Bhagat et al. (2016), for example, found that flipped instruction positively influenced high-school students' learning of mathematical concepts. Wei et al. (2020) also reported improved mathematics performance in middle-school settings, while Surjono & Mahmudi, 2023 connected flipped learning with conceptual understanding and higher-order thinking. These outcomes can be explained by the increased opportunities for active cognitive participation. When routine exposition is moved outside class, teachers may allocate more time to analysing errors, comparing solution procedures, discussing mathematical structure, and solving non-routine problems. However, this potential is realised only when classroom tasks require reasoning rather than simple reproduction. A classroom in which students merely complete additional exercises after watching a video may be temporally inverted without being pedagogically transformed.

Song and Kapur (2017) demonstrated that the design of the in-class sequence matters substantially. Their comparison of productive failure and traditional flipped arrangements suggests that students' learning can be affected by whether problem solving precedes or follows direct instruction. This finding challenges the assumption that there is a single optimal flipped sequence. Instead, the design should reflect the nature of the mathematical content, students' prior knowledge, and the intended learning processes. Cevikbas and Kaiser (2020) further argue that flipped mathematics instruction can support reform-oriented teaching when classroom time is used to promote inquiry, collaboration, explanation, and differentiated assistance. Such practices align with active-learning principles because students are required to produce, evaluate, and communicate mathematical ideas. The effectiveness of the model therefore depends on the epistemic quality of the tasks and interactions, not simply on the replacement of face-to-face lectures with videos.

2.4. Self-Regulated Learning, Preparation, and Student Agency

The pre-class phase introduces substantial self-regulatory demands. Students must plan when and where to study, maintain attention, monitor understanding, regulate effort, seek help, and decide whether they are prepared to participate in subsequent classroom activities. Self-regulated learning theory is therefore central to explaining both the potential and the limitations of flipped instruction (Zimmerman, 2002). Sun and Xie (2020) showed that students in an undergraduate mathematics course displayed distinct profiles of pre-class learning behaviour. Their findings indicate that preparation cannot be treated as a binary condition in which students either watch or do not watch a video. Learners may differ in frequency of access, timing, persistence, repetition, quiz behaviour, and achievement goals. These differences are pedagogically important because they affect the extent to which students can engage with more demanding classroom tasks.

Thai et al. (2017) likewise found that learning performance depends on the specific combination of lectures, guiding questions, and feedback. This suggests that self-regulation should not be assumed as a prerequisite that all students already possess. It should be explicitly supported through short and clearly structured resources, guiding questions, retrieval activities, progress indicators, low-stakes assessments, feedback, and opportunities to clarify difficulties. Without such scaffolding, the flipped model may transfer cognitive and organisational burdens to learners who are least prepared to manage them. Students with limited digital access, weaker prior knowledge, low academic confidence, or competing responsibilities may be disproportionately affected. Consequently, preparation rates should not be interpreted only as evidence of motivation or responsibility. They may also reflect accessibility, task clarity, resource design, perceived relevance, and the availability of support. Blau and Shamir-Inbal (2017) add an important social dimension by emphasising co-creation and co-regulation. Their redesigned flipped model indicates that learning regulation can be distributed across students, peers, teachers, and digital environments. This is particularly relevant in mathematics, where collaborative explanation can help students identify gaps in reasoning, refine language, and compare representations. Quality flipped learning should therefore cultivate both individual autonomy and collective responsibility.

2.5. TPACK and the Pedagogically Coherent Use of Technology

The Technological Pedagogical Content Knowledge framework (Mishra & Koehler, 2006) provides a useful lens for understanding teacher competence in flipped environments. Technology should not determine the educational design; instead, its selection and use should emerge from the interaction among mathematical content, pedagogical purposes, learner characteristics, and contextual constraints. A technically sophisticated resource may be instructionally weak if it presents mathematics inaccurately, increases unnecessary cognitive load, or fails to prepare students for classroom activity.

In mathematics education, technological tools can expand representational and analytic possibilities. Dynamic geometry environments, graphing tools, interactive simulations, automated feedback systems, digital whiteboards, and video annotations can help learners explore variation, connect symbolic and visual representations, and revisit explanations. Nevertheless, their educational contribution depends on how they are embedded within tasks and discourse. Hwang et al. (2015) conceptualised seamless flipped learning as an approach in which mobile technology supports continuity across learning settings. Such continuity can allow students to access materials, record questions, receive prompts, and connect pre-class preparation with classroom work. However, seamless access should not be confused with seamless learning. The latter requires pedagogical alignment so that activities completed in one phase generate information or ideas that become consequential in the next. Ruiz-Palmero et al. (2023) identified several factors affecting flipped mathematics learning in online environments, highlighting the importance of instructional organisation, teacher support, digital competence, and interaction. These findings reinforce the need for professional development that integrates technological, pedagogical, and mathematical knowledge rather than treating digital training as an isolated component.

Technology can also promote flexibility and sustainability by enabling repeated access to explanations and differentiated pathways. Karjanto and Acelajado (2022), for example, associated flipped college algebra instruction with cognitive gains and improved attitudes. Yet sustainability requires more than making digital content reusable. Resources must remain mathematically accurate, accessible, adaptable, and connected to evolving curricular and learner needs.

2.6. Learning Analytics as Evidence for Adaptive Flipped Instruction

Learning analytics can strengthen flipped instruction by making aspects of the learning process visible before, during, and after classroom interaction. Digital environments may generate data on resource access, completion, replay, response accuracy, response latency, participation, navigation sequences, help seeking, and interaction patterns. These traces can provide teachers with evidence that is difficult to obtain in conventional classroom arrangements.

Escudero-Nahón and Mercado-López (2019) found that learning analytics in flipped classrooms can support the monitoring of participation, prediction of performance, identification of behavioural patterns, and adaptation of learning activities. In mathematics education, such evidence can help teachers determine whether students encountered difficulty with a specific representation, repeatedly attempted a problem, delayed preparation, or completed resources without demonstrating understanding. However, digital traces are not quality indicators by themselves. Access counts do not demonstrate conceptual learning, and video completion does not establish that a learner understood the mathematical content. Analytics become pedagogically meaningful only when they are interpreted in relation to learning goals, task demands, prior knowledge, classroom observations, and student-produced evidence. A high-quality analytics-supported flipped design therefore links data to actionable instructional decisions. Before class, teachers may use quiz responses to identify misconceptions, organise flexible groups, select examples, or modify the lesson sequence. During class, they may combine dashboard information with questions, written work, and dialogue to determine who requires additional support. After class, they may examine patterns across activities to evaluate whether particular resources or tasks contributed to learning.

Sun and Xie (2020) latent-profile analysis illustrates the value of moving beyond aggregate measures. By identifying different behavioural patterns, analytics can reveal that students arrive at similar outcomes through different learning pathways or that comparable activity levels conceal distinct forms of engagement. Such findings support differentiated intervention, but they also require caution. Behavioural data should not be used to assign fixed deficit labels or make high-stakes predictions without contextual interpretation. Ethical considerations are therefore integral to the quality of learning analytics. Students should understand what data are collected, how they are interpreted, and how they contribute to instructional support. Data minimisation, privacy protection, transparency, and human oversight are essential. The pedagogical purpose of analytics should be formative: to expand opportunities for timely assistance and improve instructional design rather than to intensify surveillance.

2.7. Didactical Suitability as a Multidimensional Framework for Quality

Didactical suitability provides an integrative perspective for evaluating why some flipped mathematics interventions are more effective than others. Rather than defining quality through a single outcome or technological feature, this perspective considers the degree of coherence among mathematical content, learner characteristics, interaction, resources, affect, and context.

- Epistemic suitability concerns the mathematical richness, accuracy, representativeness, and coherence of the content. Pre-class materials should not reduce mathematics to isolated rules or procedures. They should present definitions, representations, connections, examples, and reasoning that are appropriate to the intended learning goals. In-class tasks should extend rather than duplicate the pre-class experience by requiring application, comparison, justification, modelling, or generalisation.
- Cognitive suitability refers to the relationship between instructional demands and learners' prior knowledge, developmental trajectories, and potential learning progression. Flipped materials that are excessively dense, linguistically complex, or insufficiently scaffolded may create barriers before students enter the classroom. The positive but heterogeneous effects reported in previous meta-analyses suggest that cognitive fit is a major source of variation across interventions (Güler et al., 2023; Haziki et al., 2025; Yakar, 2021).
- Interactional suitability concerns the quality of discourse, feedback, collaboration, and teacher responsiveness. The classroom phase should provide opportunities to explain reasoning, question assumptions, compare strategies, and receive feedback that advances understanding. The findings of Love et al. (2014), Wasserman et al. (2017), and Cevikbas and Kaiser (2020) suggest that students' experiences are shaped by how classroom interaction is redesigned after initial content delivery is moved outside class.
- Mediational suitability addresses the adequacy of technological resources, time, infrastructure, accessibility, and representational tools. A high-quality flipped design requires resources that students can access reliably and revisit as needed. It also requires sufficient classroom time for the intended mathematical practices. Technological availability alone is not enough; resources must be usable, inclusive, and aligned with pedagogical purposes.
- Affective suitability encompasses motivation, confidence, interest, agency, belonging, and willingness to participate. Kaya (2018) found that flipped learning influenced classroom engagement in middle-school mathematics, while Karjanto and Acelajado (2022) reported improved attitudes in college algebra. Nevertheless, affective responses may vary according to workload, clarity, perceived support, and students' prior experiences with mathematics and digital learning.
- Ecological suitability refers to alignment with curricula, assessment systems, institutional conditions, professional expectations, and wider social realities. Flipped interventions may be effective in experimental settings but difficult to sustain when teachers face limited planning time, inadequate infrastructure, rigid assessment requirements, or unequal student access. Ecological suitability therefore connects classroom innovation with institutional capacity and educational policy.

This multidimensional framework explains why studies using the same label can report markedly different effect sizes. The flipped classroom is not a standardised treatment but a family of designs whose components, intensities, and implementation conditions vary. Accordingly, the quality of a flipped mathematics programme should be evaluated through the coherence of its full didactical configuration rather than through the presence of videos, platforms, or pre-class activities alone.

2.8. Empirical Evidence and the Problem of Heterogeneity

The available evidence generally supports the effectiveness of flipped learning in mathematics, but it does not justify the conclusion that all implementations are equally beneficial. Güler et al. (2023) reported a positive overall effect of flipped classroom models in mathematics education, while Yakar (2021) identified favourable effects on the achievement of primary- and secondary-school students. Haziki et al. (2025) similarly reported a statistically significant medium effect in higher education mathematics. Across disciplines and educational levels, Strelan et al. (2020) also found a positive effect on student performance, while Helsa et al. (2023) reported favourable effects of hybrid learning on mathematical conceptual understanding. However, pooled estimates conceal variation associated with educational level, mathematical domain, intervention duration, research design, classroom activities, assessment instruments, and technological configuration. The comparison of these reviews suggests that the central research question should move beyond whether flipped learning works and examine under what pedagogical, technological, and institutional conditions it produces reliable benefits. Individual studies illustrate this variation. Atta and Bonyah (2023) reported strong outcomes in plane geometry among pre-service teachers, whereas other studies in calculus, linear algebra, algebra, and secondary mathematics have shown smaller or more context-dependent effects (Love et al., 2014; Wasserman et al., 2017; Wei et al., 2020). These differences may arise from the nature of the mathematical content, the baseline preparation of participants, the design of classroom tasks, or the sensitivity of outcome measures.

The distinction between achievement and broader educational quality is also important. An intervention may increase examination scores without strengthening self-regulation, mathematical discourse, transfer, or professional competence. Conversely, improvements in engagement, confidence, collaboration, and instructional reasoning may not be immediately reflected in standardised assessments. A comprehensive synthesis should therefore examine multiple outcomes and avoid reducing quality to a single pooled achievement measure.

2.9. An Integrated Model of Quality Indicators

Drawing on the theoretical and empirical literature, quality in flipped mathematics teacher education can be conceptualised as the coherence among seven interdependent dimensions:

1. Mathematical and epistemic quality, reflected in accurate content, meaningful representations, conceptual connections, and cognitively demanding tasks.
2. Pedagogical alignment, demonstrated by an explicit relationship between pre-class preparation, in-class activity, formative assessment, and intended learning outcomes.
3. Teacher professional knowledge, including pedagogical content knowledge, mathematical knowledge for teaching, technological-pedagogical competence, and the capacity to interpret student reasoning.
4. Self-regulation and learner support, reflected in structured preparation, accessible resources, guidance, feedback, accountability mechanisms, and opportunities for help seeking.
5. Interaction and active mathematical participation, including collaborative problem solving, mathematical argumentation, productive struggle, differentiated assistance, and responsive teaching.
6. Evidence-informed adaptation, in which assessment evidence and learning analytics are used to modify grouping, pacing, feedback, resource design, and task selection.
7. Equity, ethics, and ecological sustainability, encompassing digital access, accessibility, privacy, gender responsiveness, institutional support, curricular alignment, and the long-term feasibility of implementation.

These dimensions are mutually dependent. High-quality technology cannot compensate for mathematically weak content; cognitively demanding tasks may fail when preparation is inaccessible; extensive analytics may be educationally irrelevant when teachers lack time or expertise to interpret them; and strong short-term outcomes may not be sustainable without institutional support and professional reflection.

The present review therefore defines flipped classroom quality not as fidelity to a fixed instructional sequence, but as the degree of alignment among mathematical content, pedagogy, technology, assessment, analytics, equity, and teacher learning. This definition provides a theoretically grounded basis for interpreting heterogeneity across studies and for identifying indicators that can guide the design, evaluation, and improvement of professional development programmes in mathematics education.

3. Materials and Methods

This section describes the procedures used to identify, select, appraise, and synthesise the available evidence on flipped classroom implementation in mathematics education and teacher professional development. The methodological design combined a systematic review with quantitative meta-analysis and a complementary qualitative synthesis of instructional quality indicators. Particular attention was given to transparency, replicability, methodological heterogeneity, and the distinction between evidence derived from statistically comparable studies and evidence used for theoretical or contextual interpretation. The review process was organised around PRISMA 2020, while the statistical strategy was designed to account for variation in educational level, mathematical content, intervention duration, technological configuration, and study design.

3.1. Study Design and Reporting Standard

A systematic review with quantitative meta-analysis was conducted to estimate the effects of flipped classroom interventions in mathematics education and to identify evidence-based indicators of implementation quality. The study was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 statement, which guided the identification, screening, eligibility assessment, inclusion, and reporting of studies.

Before the synthesis was undertaken, the research team defined the review questions, eligibility criteria, search sources, screening stages, extraction variables, quality-assessment domains, and statistical procedures. This prior specification was intended to reduce post hoc decision-making and increase the transparency and reproducibility of the review. The methodological design comprised two complementary components. The first was a quantitative synthesis of studies reporting effect-size information that could be standardised and compared statistically. The second was a structured qualitative synthesis of intervention characteristics, teacher-development components, technological resources, assessment practices, self-regulation supports, and learning-analytics applications. This combined approach was necessary because not all relevant studies reported sufficient statistical information for inclusion in the pooled model, although they still provided evidence concerning the pedagogical and organisational conditions associated with effective flipped mathematics instruction.

The review was developed within the first phase of the FID2024-108 project, whose general purpose is to determine the effect and quality indicators of mathematics teacher professional development based on flipped classroom strategies at national and international levels. The project report documents a two-phase design consisting of meta-analysis followed by a mixed-methods stage aimed at contextualising the findings and supporting the formulation of professional-development and public-policy recommendations. Because the study used published studies, aggregated statistical results, and previously generated secondary evidence, no direct interaction with participants occurred during the systematic-review phase. Accordingly, new informed consent was not required. Nevertheless, the broader project protocol was subjected to institutional methodological review and submitted for bioethical consideration as part of the project governance process.

3.2. Information Sources and Search Strategy

The literature search was designed to identify empirical studies on flipped or inverted learning in mathematics education, with particular attention to teacher education, professional development, and measurable learning outcomes. Searches were conducted in Academic Search Ultimate, ERIC, SciELO, SAGE Journals Online, Education Full Text, Education Research Complete, and Applied Science and Technology Source Ultimate. These sources were selected because they provide broad coverage of education, mathematics teaching, instructional technology, and applied research.

Supplementary searches were carried out through Google Scholar, backward citation tracking, forward citation tracking, and the examination of relevant grey literature. Reference lists from prior reviews and meta-analyses were also inspected to identify studies that might not have been indexed consistently across databases.

The search strategy combined four conceptual blocks:

- flipped, inverted, blended, or hybrid learning;
- mathematics education, mathematics teaching, algebra, geometry, calculus, statistics, or mathematical problem solving;
- teacher education, pre-service teachers, in-service teachers, professional development, or teaching practice;
- achievement, conceptual understanding, engagement, self-regulation, higher-order thinking, or other measurable educational outcomes.

Boolean operators, truncation, and database-specific syntax were used to combine the search terms. A representative search structure was: ("flipped classroom" OR "flipped learning" OR "inverted classroom" OR "blended learning") AND ("mathematics education" OR mathematics OR algebra OR geometry OR calculus OR statistics) AND ("teacher education" OR "professional development" OR teaching OR learning) AND (achievement OR performance OR understanding OR engagement).

The principal search window covered publications from 2013 to 2023 in English or Spanish. This period was selected because it corresponds to the consolidation of flipped learning as a recognisable educational model and to the expansion of digital learning environments in mathematics education. Studies published after 2023 were not incorporated retrospectively into the primary search set but were used to contextualise the discussion and compare the review findings with more recent evidence. According to the project progress report, 500 records were initially retrieved from the indexed databases, with a further 80 records identified through Google Scholar, citation tracking, and other sources. After removing 102 duplicates, 478 records remained for title and abstract screening.

3.3. Eligibility, Screening, and Data Extraction

Studies were eligible for inclusion when they met the following criteria:

1. They examined an identifiable flipped, inverted, or hybrid instructional intervention in mathematics education.
2. They included a comparison-group, quasi-experimental, experimental, or interpretable pre-post design.
3. They reported quantitative outcomes related to mathematics achievement, conceptual understanding, higher-order thinking, engagement, teacher competence, or another measurable educational variable.
4. They provided sufficient methodological and statistical information to permit appraisal and, when applicable, effect-size calculation.
5. They were published in English or Spanish within the predefined search period.
6. They involved learners, pre-service teachers, in-service teachers, or professional-development participants in contexts directly related to mathematics teaching and learning.

Studies were excluded when mathematics was not the principal disciplinary domain, the intervention could not be distinguished from a broader blended-learning programme, the outcomes were not quantitatively interpretable, the design did not permit a meaningful effect estimate, or the report duplicated data published elsewhere. Conceptual papers, editorials, opinion articles, and reviews were retained only as contextual references and were not entered into the quantitative synthesis. Screening was conducted in two stages. First, titles and abstracts were examined for thematic relevance. Second, potentially eligible reports were assessed in full text against the inclusion and exclusion criteria. The progress report indicates that 478 titles and abstracts were screened and that 312 records were excluded because they did not explicitly address flipped learning and mathematics education. A total of 166 reports proceeded to full-text eligibility assessment. Of these, 42 were excluded because they did not report measurable student outcomes, 29 because their design did not meet the review requirements or did not identify an active teaching sample, and 25 because they focused on disciplines other than mathematics. Seventy studies were retained in the broader quantitative and qualitative evidence base.

A standardized data extraction matrix was developed to ensure consistency and methodological rigor across the included studies. The matrix captured comprehensive information on each study, including the author(s) and publication year, country and educational context, educational level, sample characteristics and size, study design, mathematical content domain, duration and intensity of the intervention, characteristics of the pre-class and in-class phases of the flipped classroom model, technological tools and digital resources employed, comparison condition, outcome variables and measurement instruments, descriptive and inferential statistical data (including means, standard deviations, standard errors, confidence intervals, and sample sizes), teacher professional development components, formative assessment procedures, self-regulation supports, learning analytics indicators, gender-related information, and implementation constraints and contextual factors. This comprehensive extraction process facilitated the systematic synthesis of methodological characteristics, intervention features, and outcome evidence, while enabling subsequent qualitative synthesis and quantitative meta-analytic analyses.

Where multiple outcomes were reported, priority was given to mathematics achievement and conceptual understanding for the main quantitative synthesis. Additional outcomes were retained for supplementary interpretation and the development of quality indicators.

3.4. Quantitative Synthesis

The primary quantitative outcome was the standardised mean difference, expressed as Cohen's d. This metric was selected because the included studies used different achievement tests, scales, course assessments, and locally developed instruments. Standardisation allowed results to be compared across educational levels and mathematical domains.d

For studies reporting group means, standard deviations, and sample sizes, the effect size was calculated as the difference between the intervention and comparison means divided by the pooled standard deviation. When standard errors were available, 95% confidence intervals were calculated as d ± 1.96(SE). Positive effect sizes indicated better outcomes for participants exposed to flipped instruction, whereas negative values favoured the conventional or comparison condition.

$$CI_{95\%} = d \pm 1.96(SE).$$

A random-effects model was specified as a priori because the studies were not assumed to estimate a single common intervention effect. The interventions differed in educational level, country, mathematical content, sample composition, technological platform, duration, teacher preparation, classroom design, and outcome measurement. Under these conditions, a random-effects model provides a more defensible estimate of the average effect across a distribution of true effects. The 19-study synthesis used inverse-variance random-effects weights. Each study's contribution to the pooled estimate was therefore determined by its statistical precision while also incorporating between-study variance. The pooled result was interpreted together with its confidence interval and heterogeneity statistics rather than as a universal estimate applicable to every context.

Heterogeneity was assessed using Cochran's Q, the I^2 statistic, and τ^2. Cochran's Q tested whether the variability among observed effects exceeded that expected from sampling error. I^2 estimated the proportion of total variability attributable to genuine between-study differences, while τ^2 represented the estimated variance of the distribution of true effects. High heterogeneity was interpreted as evidence that intervention effectiveness depended on contextual and pedagogical moderators rather than as evidence that the pooled estimate lacked value.$QI^2\tau^2QI^2\tau^2$

A complementary evidence set comprising 46 effect sizes from Yakar (2021) meta-analysis of mathematics achievement in primary and secondary education was retained for comparative validation. This evidence set was not statistically combined with the 19-study model because study-level standard errors and variance estimates were not uniformly recoverable from the published source. Combining the two datasets would therefore have required assigning incomplete or non-equivalent precision estimates and could also have introduced dependency through the possible inclusion of overlapping primary studies. Instead, the published pooled estimate and heterogeneity statistics reported by Yakar (2021) were compared with the results of the independently reconstructed 19-study random-effects model. Convergence between the two pooled estimates was interpreted as supportive evidence regarding the general direction and approximate magnitude of the effect, but not as an independent replication or as justification for producing a single combined estimate.

Potential moderators were considered in relation to educational level, mathematical domain, intervention duration, study design, technological configuration, teacher preparation, and the alignment between pre-class resources and in-class mathematical activity. However, formal subgroup analysis and meta-regression were restricted by the limited number of statistically comparable studies within each category and by incomplete reporting of study-level variance and intervention characteristics. Moderator patterns were therefore interpreted descriptively and theoretically rather than through underpowered inferential models. This decision reduced the risk of unstable coefficients, false-positive moderator effects, and overinterpretation of differences based on small subgroups.

3.5. Quality Assessment and Risk-of-Bias Logic

Methodological quality was assessed using a domain-based approach rather than a single composite numerical score, acknowledging that overall quality indices may obscure important methodological strengths and weaknesses by assigning arbitrary weights to distinct dimensions of study quality. The appraisal therefore examined multiple domains relevant to the internal and external validity of the included studies, including the clarity of the research design, comparability of intervention and control groups, transparency of participant selection procedures, validity and reliability of outcome measures, completeness of attrition reporting, treatment of missing data, fidelity of flipped-classroom implementation, equivalence of instructional time between comparison conditions, control of potential confounding variables, appropriateness of statistical analyses, completeness of effect-size reporting, and the risk of selective outcome reporting. This multidimensional assessment provided a more nuanced evaluation of methodological rigor and informed the interpretation of the evidence synthesized in both the qualitative review and the meta-analysis.

Particular attention was given to quasi-experimental studies because baseline differences, teacher assignment, self-selection, and unequal access to technology could influence outcomes independently of the intervention. Teacher effects were also considered important, as improvements attributed to flipped learning may partly reflect differences in instructor experience, pedagogical competence, feedback practices, or familiarity with educational technology. Locally developed achievement tests were not automatically considered invalid, but they were examined in relation to the reporting of reliability, content alignment, scoring procedures, and evidence of validity. Studies relying exclusively on course grades or teacher-generated assessments were interpreted more cautiously when instrument quality was insufficiently documented. Implementation fidelity was assessed by examining whether the study described both the pre-class and in-class phases, whether students had meaningful access to the digital materials, whether pre-class preparation was monitored, and whether classroom time was reorganised toward active mathematical activity. Studies that merely replaced face-to-face lectures with videos without redesigning classroom interaction were considered weaker representations of the flipped-learning model.

Rather than excluding all studies with methodological limitations, the review incorporated risk-of-bias judgements into the interpretation of certainty. Findings supported by transparent designs, comparable groups, validated measures, and complete statistical reporting were given greater interpretive weight than results derived from poorly described or highly confounded designs. Risk-of-bias judgements were completed independently by two members of the review team using the predefined domains. Each domain was classified as low risk, some concerns, or high risk. Disagreements were resolved through discussion and re-examination of the full-text report; when consensus could not initially be reached, a third reviewer adjudicated the decision. The assessment was conducted at the study level rather than converted into a single composite score, because weaknesses in participant allocation, outcome measurement, attrition, implementation fidelity, and selective reporting have different implications for internal validity. The resulting domain-level judgements were summarised descriptively and incorporated into the interpretation of the pooled evidence.

3.6. Evidence-to-Indicator Synthesis

The second analytical component focused on identifying evidence-based quality indicators for flipped mathematics instruction and teacher professional development. These indicators were generated through structured thematic coding of intervention characteristics, explanatory findings, reported implementation conditions, and methodological conclusions.

Candidate quality indicators were retained only when they satisfied four complementary criteria: they were consistently identified across multiple empirical studies or systematic reviews, demonstrated a clear theoretical association with mathematics learning, teacher expertise, or instructional quality, could be directly observed or documented during implementation, and were amenable to improvement through instructional design, professional development, institutional support, or educational policy. The coding process synthesized evidence related to the quality of mathematical tasks, alignment between pre-class and in-class learning activities, teacher feedback, formative assessment, learner preparation, digital accessibility, self-regulation, collaborative problem solving, teacher reflection, and the use of learning analytics. These findings were subsequently organized according to five complementary theoretical perspectives Pedagogical Content Knowledge (PCK), Mathematical Knowledge for Teaching (MKT), Technological Pedagogical Content Knowledge (TPACK), Self-Regulated Learning (SRL), and the Didactical Suitability framework to provide a robust conceptual foundation for the proposed indicators. Within this framework, learning analytics was conceptualized as a cross-cutting evidence mechanism rather than an independent technological dimension, reflecting its pedagogical value in supporting data-informed instructional decisions related to grouping, pacing, formative feedback, and task selection. Based on the convergence of empirical evidence and theoretical foundations, the resulting framework organizes the quality indicators into seven interrelated dimensions: (i) mathematical and epistemic quality, (ii) alignment between pre-class preparation and classroom activities, (iii) teacher professional knowledge, (iv) learner self-regulation and instructional support, (v) interaction and active mathematical participation, (vi) evidence-informed instructional adaptation, and (vii) equity, ethics, and ecological sustainability.

This synthesis made it possible to move beyond a binary judgement of whether flipped learning was effective. Instead, it identified the conditions that appeared to explain why some interventions produced larger or more consistent effects than others. The resulting indicators were therefore conceived as tools for designing, evaluating, and improving mathematics teacher professional-development programmes rather than as a prescriptive checklist detached from context.

4. Results

This section presents the results of the systematic selection process, the quantitative synthesis of effect sizes, the assessment of between-study heterogeneity, and the thematic identification of quality indicators for flipped mathematics education. The findings are organised to address two complementary purposes: first, to estimate the average effect of flipped classroom interventions on mathematics learning; and second, to determine which pedagogical, technological, and professional-development conditions may explain the variability observed across studies. Because the available evidence differed in statistical completeness, the results distinguish between the 19 studies for which standard errors or equivalent uncertainty estimates could be recovered and the complementary set of 46 primary- and secondary-level effect sizes reported by Yakar (2021).

4.1. Study Selection and Evidence Base

The search process identified 580 records potentially related to flipped learning, mathematics education, teacher professional development, and measurable educational outcomes. Of these, 500 records were retrieved from the selected bibliographic databases, while 80 additional records were located through Google Scholar, backward and forward citation tracking, and complementary sources. After the removal of 102 duplicate records, 478 unique titles and abstracts remained for screening. At the first screening stage, 312 records were excluded because they did not address flipped or inverted learning in mathematics education with sufficient specificity or did not report an empirical intervention relevant to the review questions. The remaining 166 reports were assessed in full text. Ninety-six reports were excluded at this stage: 42 did not provide measurable learning or teaching outcomes, 29 did not meet the required comparative-design or active-teaching-sample criterion, and 25 focused on disciplinary areas other than mathematics.

The final qualitative evidence base consisted of 70 studies. These studies informed the descriptive synthesis, the examination of intervention characteristics, and the construction of the evidence-based quality-indicator framework. The 70 included reports represented approximately 12.1% of the 580 initially identified records, illustrating the extent to which apparently relevant publications were removed after methodological and disciplinary eligibility criteria were applied. The project progress report documents these selection totals and the associated reasons for exclusion. The quantitative evidence was subsequently organised according to the completeness of the statistical information. Nineteen studies provided recoverable standard errors or equivalent information sufficient to calculate study-level confidence intervals and inverse-variance weights. These studies constituted the independent random-effects meta-analysis.

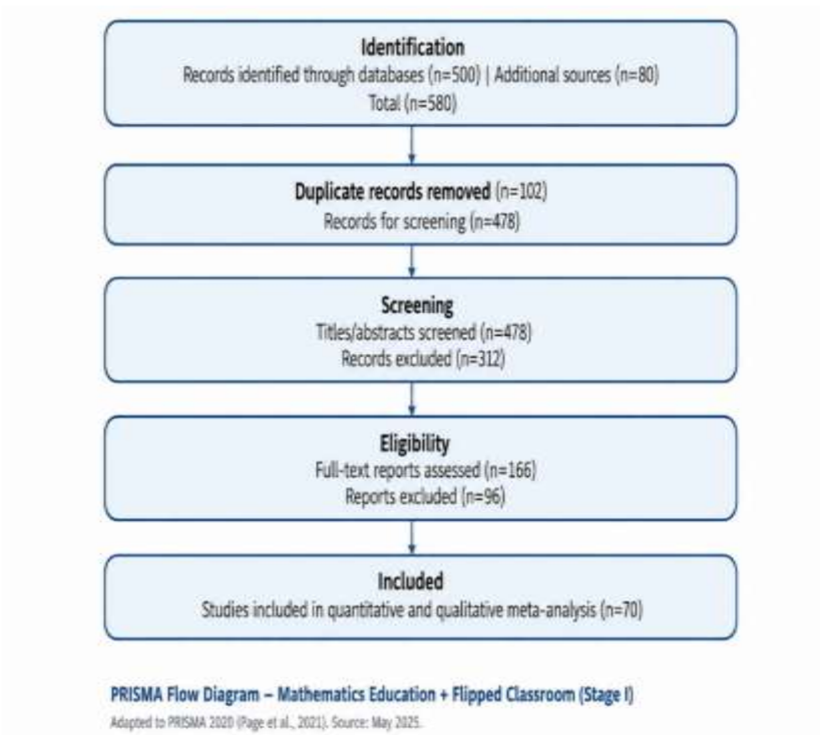

Figure 1. PRISMA 2020 flow diagram for study identification, screening, eligibility assessment, and inclusion.

A second evidence set consisted of 46 effects reported in Yakar (2021) meta-analysis of primary and secondary mathematics achievement. Because the uncertainty estimates for all individual effects were not uniformly recoverable, this dataset was retained as an external convergent synthesis rather than statistically combined with the 19-study model. This distinction is methodologically important. The 70-study qualitative corpus addressed the broader question of which characteristics define high-quality flipped mathematics education, whereas the 19-study statistical subset addressed the narrower question of the magnitude and precision of the average intervention effect. The 46-effect Yakar dataset served as an independent comparison across school levels.

The study selection process is presented following the PRISMA 2020 statement, providing a transparent and reproducible account of the identification, screening, eligibility assessment, and final inclusion of studies. The flow diagram summarizes the number of records retrieved from the selected databases, the removal of duplicate records, the screening of titles and abstracts, the full-text assessment of potentially eligible studies, the reasons for exclusion at each stage, and the final set of studies included in the qualitative synthesis and quantitative meta-analysis. This standardized reporting framework enhances methodological transparency, facilitates the reproducibility of the review process, and enables readers to evaluate the comprehensiveness and rigor of the evidence selection procedure.

Table 1. PRISMA study-flow summary

Stage	Records	Interpretive role
Records identified	580	Broad multilingual and multidisciplinary evidence pool
Records remaining after duplicate removal	478	Unique records submitted to title and abstract screening
Full-text reports assessed	166	Reports evaluated against methodological and disciplinary criteria
Studies included in qualitative synthesis	70	Evidence base for intervention characteristics and quality indicators
Studies with recoverable standard errors	19	Independent inverse-variance random-effects meta-analysis
Primary and secondary effect estimates	46	External convergent synthesis based on Yakar (2021)

The study-flow results reveal a substantial difference between the breadth of the initial literature and the smaller body of evidence suitable for rigorous synthesis. Many publications discussed flipped instruction descriptively or examined technology-enhanced teaching without providing an identifiable comparison, measurable mathematics outcome, or sufficient statistical information. Consequently, the final evidence base should be interpreted as a methodologically filtered corpus rather than a comprehensive count of all publications mentioning flipped learning.

4.2. Overall Effects and Between-Study Heterogeneity

The random-effects model based on the 19 studies with recoverable uncertainty estimates produced a pooled standardised mean difference of d = 0.522, with a standard error of 0.147 and a 95% confidence interval from 0.233 to 0.811. The overall effect was statistically significant (p = .0004) and indicated a moderate average advantage for flipped mathematics interventions over conventional or comparison conditions.$d = 0.522$ $p = .0004$

The exclusion of zero from the confidence interval supports the conclusion that flipped mathematics instruction was beneficial on average. Nevertheless, the magnitude of the pooled result should not be interpreted as a fixed expected improvement for every educational setting. The model estimated the centre of a distribution of true effects generated under substantially different conditions, including variations in educational level, mathematics domain, intervention length, digital infrastructure, teacher preparation, student characteristics, and classroom activity.

The heterogeneity statistics confirmed this variability. Cochran's test was statistically significant, Q(18) = 208.089, p < .001, indicating that the dispersion of observed effects was greater than would be expected from sampling error alone. The I^2 value of 91.35% showed that a very high proportion of observed variability was attributable to genuine between-study differences. The estimated between-study variance was τ^2 = 0.369.$Q(18) = 208.089$$p < .001$$I^2$$\tau^2 = 0.369$

These results establish that the pooled effect must be interpreted in conjunction with implementation conditions. The intervention category "flipped classroom" did not represent a uniform treatment. Across the included studies, the model ranged from the use of recorded lectures followed by routine exercises to more elaborate designs involving diagnostic quizzes, collaborative problem solving, teacher feedback, differentiated instruction, dynamic mathematical representations, and structured self-regulation supports.

The individual effect estimates reflected this diversity. Atta and Bonyah (2023) produced the largest observed effect (d = 2.59) in a plane-geometry intervention involving pre-service teachers. Such a result indicates a substantial difference between groups but should be interpreted in relation to the local design, sample, outcome instrument, and study precision rather than treated as a typical estimate. At the opposite end of the distribution, several studies showed negative or near-zero effects. Other study-level confidence intervals crossed the null value, indicating that their available estimates were insufficiently precise to exclude no difference between flipped and comparison conditions.$d = 2.59$

By contrast, several studies reported confidence intervals located entirely above the null value and therefore provided stronger evidence of benefit within their respective contexts. The coexistence of clearly positive, statistically uncertain, and negative effects reinforces the conclusion that implementation quality and contextual fit are likely moderators of effectiveness.

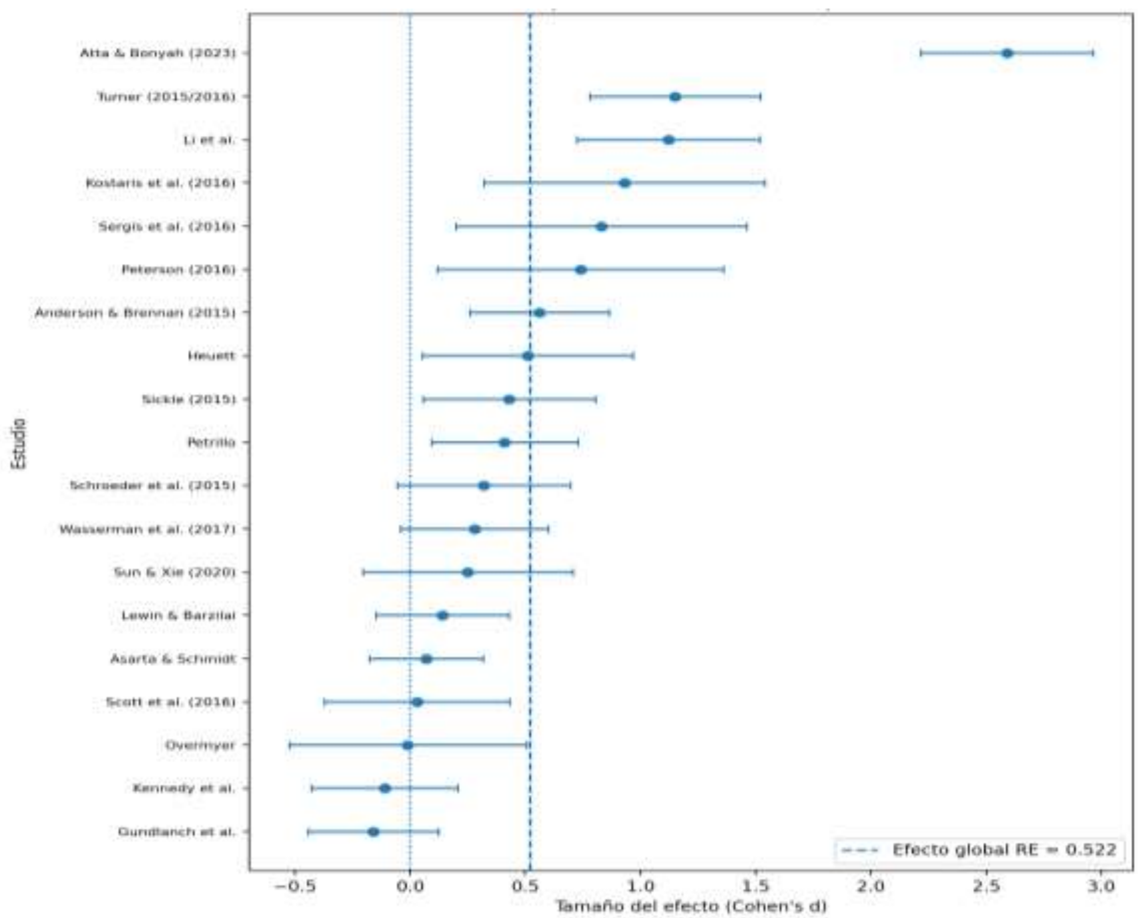

Figure 2. Random-effects forest plot for the 19 studies with recoverable standard errors. Points represent study-level Cohen's d estimates and horizontal lines represent 95% confidence intervals. The dashed vertical line marks the pooled random-effects estimate (d = 0.522), while the dotted vertical line marks the null effect (d = 0).$dd = 0.522$ $d = 0$

Table 2. Complementary random-effects evidence

Evidence set	k	Pooled d	95% CI	Q	I^2	Interpretation
Studies with recoverable standard errors	19	0.522	[0.233, 0.811]	208.089	91.35%	Moderate average benefit with very high heterogeneity

Evidence set	k	Pooled d	95% CI	Q	I^2	Interpretation
Yakar primary/secondary synthesis	46	0.510	[0.340, 0.690]	372.06	87.91%	Comparable average benefit with high heterogeneity

The similarity between the two pooled estimates is notable. Despite differences in educational level, study composition, and statistical source, both syntheses produced an average effect close to d = 0.51. This convergence increases confidence in the general direction and approximate magnitude of the association between flipped instruction and mathematics achievement. However, the high heterogeneity in both evidence sets prevents the pooled mean from being treated as a universal intervention parameter.$d = 0.51$

4.3. Convergence with Evidence from Primary and Secondary Education

The independent synthesis of 46 effect sizes reported by Yakar (2021) yielded a pooled effect of d = 0.51, with a standard error of 0.09 and a 95% confidence interval from 0.34 to 0.69. The result indicated a statistically meaningful and moderate average advantage for flipped learning in primary and secondary mathematics education.$d = 0.51$; The heterogeneity was again substantial. Cochran's statistic was Q = 372.06, and the I^2 value was 87.91%. These results indicate that most of the observed dispersion arose from systematic differences among studies rather than from sampling variation alone. Therefore, the primary- and secondary-level evidence reproduced the same central pattern observed in the 19-study model: a positive average effect accompanied by pronounced contextual variation. $Q = 372.06 I^2$

The 46 individual effects ranged from d = -0.56 to d = 3.73. Thirty-seven values were positive and nine were negative. The predominance of positive effects supports the overall direction of the pooled estimate, but the wide range demonstrates that the effectiveness of flipped mathematics instruction was not uniform. The largest estimates may reflect highly intensive or contextually favourable interventions, small-study effects, particular outcome measures, or stronger contrasts with conventional teaching. Conversely, negative effects may reflect implementation difficulties, limited student preparation, insufficient technological access, weak alignment between phases, or comparison conditions that were already pedagogically strong.-0.56

The close agreement between the 19-study estimate (d = 0.522) and the Yakar estimate (d = 0.510) provides convergent evidence across educational levels. It suggests that the average effect does not depend exclusively on whether participants were school students, university students, or pre-service teachers. Instead, the heterogeneity implies that more decisive factors may include the quality of mathematical tasks, the connection between pre-class and in-class activity, scaffolding of self-regulation, teacher professional knowledge, and the availability of timely formative feedback. This convergence should nevertheless be interpreted cautiously because the two evidence sets were not constructed under identical inclusion and coding procedures and may contain overlapping primary studies. The Yakar dataset was therefore used for comparative validation rather than pooled directly with the 19-study model. $d = 0.522 d = 0.510$; The pattern displayed in Figure 3 provides an important visual complement to the pooled estimate because it shows that the average effect of d = 0.51 is not produced by a tightly clustered set of homogeneous studies. Most estimates are located to the right of the null line, confirming that positive effects predominate across primary and secondary mathematics contexts. However, the marked asymmetry and extended upper tail indicate that a small number of studies reported exceptionally large effects, while several estimates remained close to zero or negative. This distribution suggests that the overall benefit of flipped learning is robust in direction but highly variable in magnitude.

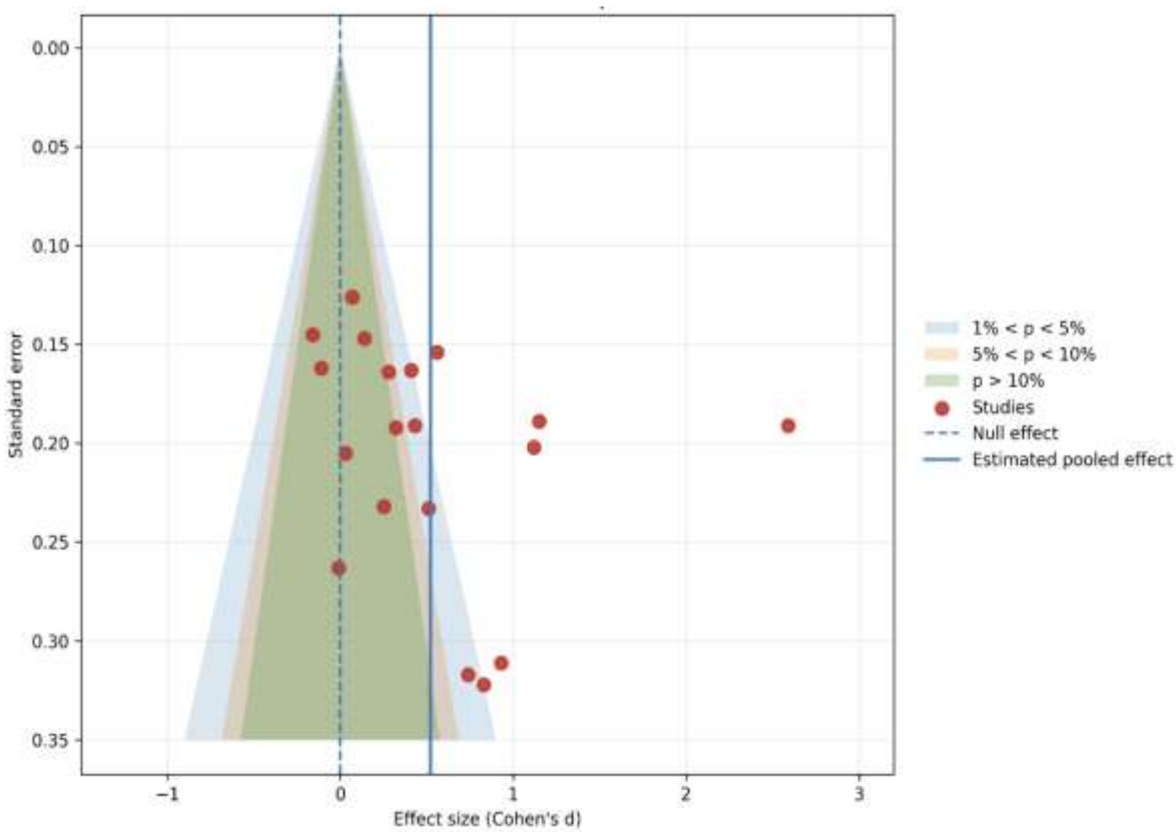

Figure 3. Contour-enhanced funnel plot of the 19 studies included in the random-effects meta-analysis.

Figure 3 presents a contour-enhanced funnel plot for the 19 studies included in the random-effects meta-analysis. The visual pattern suggests some asymmetry, with several studies concentrated on the positive side of the pooled estimate and one notably extreme positive effect. However, this apparent asymmetry should not be interpreted as conclusive evidence of publication bias because the meta-analysis includes a limited number of studies and exhibits very high heterogeneity. In this context, the funnel plot is better interpreted as an exploratory indicator of dispersion, study influence, and possible small-study effects than as a definitive diagnostic of selective publication.

The figure therefore reinforces the interpretation that the pooled effect represents a central tendency across diverse implementations rather than a uniform intervention response. In substantive terms, the dispersion is consistent with differences in mathematical content, intervention intensity, teacher preparation, technological support, student readiness, and the degree of alignment between pre-class preparation and in-class mathematical activity. Accordingly, Figure 3 supports the need to interpret effectiveness through implementation quality and contextual fit, rather than through the flipped-classroom label alone.$d = 0.51$

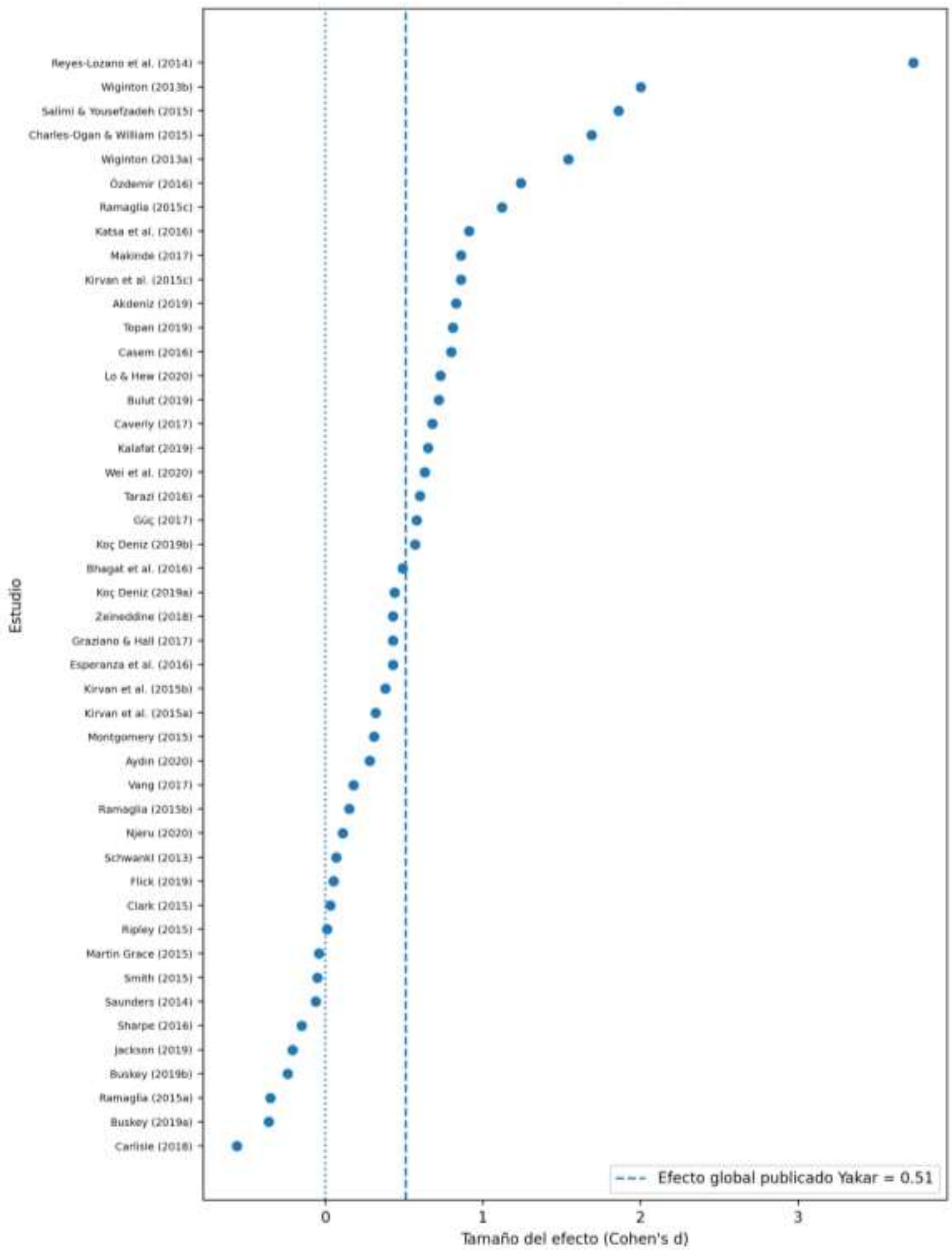

Figure 4. Distribution of 46 effect sizes in primary and secondary mathematics education reported by Yakar (2021). The dashed line represents the published pooled effect (d = 0.51), while the dotted line represents the null effect (d = 0). The distribution shows a positive central tendency accompanied by substantial between-study dispersion. $d = 0.51$

4.4. Evidence-Based Quality Indicators for Flipped Mathematics Education

The thematic synthesis of the 70-study evidence base identified seven interdependent indicators associated with high-quality flipped mathematics instruction and professional teacher development. These indicators were not derived from the frequency of isolated technological features. Instead, they emerged from recurring relationships among instructional design, teacher knowledge, student activity, assessment, access, and evidence-informed adaptation.

4.4.1. Intentional pre-class design

The first indicator concerns the pedagogical quality of the pre-class phase. Effective preparation was characterised by concise and accessible resources, explicit mathematical objectives, manageable cognitive demand, guiding questions, and mechanisms that enabled teachers to determine whether students had engaged with and understood the material.

The evidence suggests that the pre-class phase was most productive when it introduced foundational ideas without attempting to replace the complete learning process. Short explanations, worked examples, visual representations, retrieval questions, and low-stakes checks helped students develop sufficient initial familiarity to participate in subsequent classroom activity. The indicator therefore extends beyond video availability. A technically polished recording may remain ineffective when it is too long, disconnected from the lesson, inaccessible to some students, or focused on procedural imitation. Observable evidence of intentional design includes alignment between resources and learning outcomes, clear instructions, accessibility across devices, guiding prompts, and evidence of completion or understanding. For teacher professional development, this indicator requires competence in content curation, multimedia design, cognitive-load management, accessibility, and anticipation of likely student difficulties.

4.4.2. Mathematically rich in-class activity

The second indicator is the degree to which classroom time is reorganised around cognitively demanding mathematics. The synthesis showed that the educational value of inversion depended primarily on what occurred after students had completed the preparatory phase.

High-quality implementations used classroom time for problem solving, comparison of strategies, interpretation of representations, mathematical argumentation, modelling, error analysis, conjecture, and teacher-guided discussion. These activities required learners to use and reorganise knowledge rather than merely reproduce procedures introduced before class. This indicator differentiates pedagogically transformed flipped learning from simple temporal relocation. When classroom time is devoted to repetition of the displaced lecture or completion of routine exercises without feedback, the instructional model remains largely teacher-centred despite the use of digital resources. The professional-development implication is that teachers require preparation in task design, orchestration of discourse, anticipation of strategies, selection of student work, and management of productive struggle.

4.4.3. Formative assessment and responsive feedback

The third indicator concerns the connection between evidence generated before class and instructional decisions made during class. Effective interventions used diagnostic quizzes, submitted questions, short written responses, digital traces, or preliminary tasks to identify misconceptions and determine the level of preparation.

The strongest designs did not collect this evidence merely for accountability. Teachers used it to modify pacing, select examples, organise groups, initiate discussion, and provide targeted feedback. In this way, formative assessment served as the bridge between asynchronous preparation and synchronous mathematical learning. Observable evidence includes diagnostic checks, flexible grouping, feedback connected to specific reasoning, opportunities for revision, and instructional changes based on student responses. Professional development should therefore include assessment literacy, interpretation of misconceptions, feedback design, and the use of evidence to support immediate pedagogical adaptation.

4.4.4. Integrated mathematics, pedagogy, and technology knowledge

The fourth indicator is the teacher's capacity to integrate disciplinary, pedagogical, and technological knowledge. The variability of effect sizes showed that technology alone did not account for stronger outcomes. Effective implementation required teachers to select tools that supported particular mathematical representations, anticipate how students might interpret those representations, and design transitions between independent study and collective reasoning. This integrated expertise corresponds to the interaction among pedagogical content knowledge, mathematical knowledge for teaching, and TPACK. It becomes visible in the quality of explanations, examples, questions, digital resource choices, and responses to student reasoning. Professional-development programmes should therefore avoid treating platform use or video production as sufficient preparation. More effective approaches include video-based reflection, Lesson Study, collaborative task analysis, coached implementation, and iterative redesign based on classroom evidence.

4.4.5. Explicit support for self-regulated learning

The fifth indicator is the presence of structured support for planning, monitoring, persistence, retrieval, and help seeking. The evidence indicated that successful preparation could not be assumed simply because digital resources were available. Students differed in their capacity to manage time, sustain attention, identify misunderstanding, and seek assistance. High-quality interventions therefore included pacing guidance, preparation schedules, embedded questions, progress indicators, retrieval opportunities, reminders, and clear routes for requesting support. This indicator is particularly relevant to equity because students with weaker prior knowledge, limited digital experience, or competing responsibilities may experience greater difficulty in managing asynchronous preparation. Professional development should prepare teachers to design scaffolds and gradually transfer responsibility to learners rather than presuming fully developed autonomy from the outset.

4.4.6. Equitable access and inclusive design

The sixth indicator concerns access to technology, time, language, accessibility features, and alternative modes of participation. The review showed that flipped learning may expand flexibility while simultaneously creating new barriers when participation depends on stable connectivity, personal devices, quiet study spaces, or advanced digital competence. Quality therefore requires low-bandwidth alternatives, downloadable resources, captions, readable formats, flexible deadlines, device-access options, and non-digital alternatives when necessary. Equity should be treated as a design condition rather than as a post-implementation adjustment. The progress report includes exploratory country, gender, and technology-use graphics on pages 4 and 5. These images may be useful for internal project planning, but they should not be included in the scientific results section unless the variables, samples, coding procedures, and raw data used to generate them are fully documented. The report itself presents these visualizations as comparative project analyses rather than as validated moderator results.

4.4.7. Analytics-informed pedagogical adaptation

The seventh indicator is the responsible use of learning analytics to support instructional decisions. Digital traces such as access frequency, completion patterns, quiz responses, response latency, and help-seeking behavior can provide evidence of students' preparation and difficulty. However, the synthesis indicated that activity counts are not direct measures of learning quality. Analytics become educationally meaningful only when they are interpreted alongside mathematical goals, student work, classroom observation, and contextual knowledge. A high number of video views, for example, may indicate engagement, confusion, or technical difficulty. Observable evidence of this indicator includes the use of learning traces to adapt pacing, organize groups, identify misconceptions, select tasks, and provide support. Professional development should therefore include data literacy, interpretive caution, privacy protection, transparency, and avoidance of deficit-based classifications.

Table 3. Quality indicators for flipped mathematics teacher education

Quality indicator	Observable evidence	Professional-development focus
Intentional pre-class design	Goal alignment, concise resources, guiding questions, accessibility, and completion evidence	Content curation, multimedia design, and cognitive-load management
Mathematically rich in-class work	Reasoning, multiple representations, discourse, modelling, and non-routine tasks	Task design, mathematical discussion, and classroom orchestration
Formative assessment and feedback	Diagnostic checks, responsive grouping, feedback, and opportunities for revision	Assessment literacy and interpretation of student thinking
Integrated teacher knowledge	PCK, MKT, and TPACK reflected in explanations, examples, and technology selection	Lesson Study, video analysis, coached implementation, and collaborative redesign
Self-regulation support	Planning prompts, pacing guidance, retrieval activities, progress monitoring, and help seeking	Scaffold design and gradual release of responsibility
Equitable digital access	Low-bandwidth formats, captions, alternative devices, flexible access, and inclusive resources	Universal design, accessibility, and equity-sensitive planning
Analytics-informed adaptation	Learning traces interpreted in relation to mathematical evidence and instructional action	Data literacy, privacy, ethical interpretation, and responsive teaching

Taken together, these seven indicators show that the effectiveness of flipped mathematics education is produced by the coherence of the instructional system rather than by any single resource or platform. The moderate pooled effects indicate that the model has educational potential, while the very high heterogeneity demonstrates that outcomes depend on whether its components are aligned with mathematical goals, learner needs, teacher expertise, and institutional conditions.

5. Discussion

This study examined whether flipped classroom interventions improve mathematics learning and which evidence-based quality indicators should guide their incorporation into mathematics teacher professional development. The findings provide convergent support for a moderate positive effect while simultaneously demonstrating that flipped learning cannot be treated as a standardized intervention. The pooled estimate obtained from the 19 studies with recoverable uncertainty information, $d = 0.522$, was closely aligned with the independent estimate reported by Yakar (2021), $d = 0.510$. This convergence across different educational levels and evidence strengthens confidence in the overall direction of the effect. Nevertheless, the very high heterogeneity observed in both syntheses indicates that effectiveness depends on the pedagogical, technological, institutional, and professional conditions under which the model is implemented.

The principal contribution of the study is therefore not limited to confirming that flipped mathematics instruction can produce favorable outcomes. More importantly, it identifies the conditions that distinguish pedagogically substantive implementations from interventions based primarily on the relocation of exposition. The synthesis demonstrates that intentional pre-class design, mathematically rich classroom activity, formative assessment, integrated teacher knowledge, support for self-regulation, equitable access, and analytics-informed adaptation operate as an interconnected system. These dimensions provide a theoretically grounded explanation for the considerable variation observed among individual study effects.

5.1. Interpretation of RQ1: Effects of Flipped Classroom Interventions on Mathematics Learning

The first research question examined the extent to which flipped classroom interventions improve mathematics learning. The pooled effect of d = 0.522 represents a moderate educational advantage for flipped instruction over conventional or comparison approaches. This result is consistent with the broader international evidence base. Strelan et al. (2020) reported a positive overall effect of flipped classroom instruction across disciplines and educational levels, while Güler et al. (2023) found statistically significant benefits specifically within mathematics education. Yakar (2021) similarly identified a moderate positive effect among primary- and secondary-school learners, and Haziki et al. (2025) reported a significant pooled effect of in higher education mathematics.$d = 0.522d = 0.474$

The proximity of these estimates is scientifically relevant. The current pooled effect of d = 0.522 is slightly higher than the estimate reported by Haziki et al. (2025), but remains within the same moderate-effect range. This consistency across school, university, and teacher-education settings suggests that flipped learning has a broadly reproducible potential to improve mathematics performance. However, the results do not support a deterministic interpretation according to which any flipped intervention will outperform conventional teaching. Rather, flipped instruction appears to create an opportunity structure within which improved learning becomes more likely when the model is supported by rigorous instructional design.$d = 0.522$

The high heterogeneity observed in the present synthesis, $I^2 = 91.35\%$, is also consistent with previous meta-analyses. Yakar (2021) reported $I^2 = 87.91\%$, while the wider literature has repeatedly shown that flipped-learning effects vary according to educational level, subject content, intervention duration, instructional sequencing, assessment instruments, and implementation fidelity (Güler et al., 2023; Strelan et al., 2020). Consequently, the pooled mean should be interpreted as the centre of a distribution of context-dependent effects rather than as a fixed expected gain.

This interpretation is reinforced by the study-level results. The very large effect reported by Atta and Bonyah (2023) in plane geometry among pre-service teachers suggests that flipped learning may be particularly effective when content, instructional activities, and assessment are tightly aligned. Their intervention required participants to engage with mathematical explanations before class and subsequently apply those ideas through guided and collaborative activity. The strong effect may therefore reflect a coherent didactical configuration in which pre-class preparation directly enabled higher-order classroom participation.

At the same time, very large estimates warrant methodological caution. They may be influenced by small samples, locally developed outcome measures, researcher involvement in implementation, or weak control conditions. The current findings therefore support the educational relevance of such interventions without assuming that their observed magnitude will generalize automatically to other contexts.

The smaller and statistically uncertain estimates found in studies such as Wasserman et al. (2017), Nurhajarurahmah et al. (2026), and Sun and Xie (2020) demonstrate that the effectiveness of flipped learning is conditional. Null or imprecise findings do not necessarily indicate that the pedagogical model is ineffective. They may instead reflect inconsistent pre-class participation, inadequate scaffolding, insufficient distinction between pre-class and in-class activity, limited intervention duration, or outcome instruments that did not adequately capture the targeted conceptual development.

This explanation is consistent with Cevikbas and Kaiser (2020) reform-oriented interpretation of flipped mathematics education. From this perspective, the active ingredient is not the use of videos or digital platforms, but the reorganization of mathematical activity toward inquiry, explanation, collaboration, feedback, and differentiated support. A course may be structurally inverted while remaining pedagogically conventional if classroom time continues to be dominated by routine exercises, teacher explanation, or procedural reproduction.

Evidence from Bhagat et al. (2016), Surjono & Mahmudi, 2023, and Wei et al. (2020) further supports the importance of conceptual and cognitive activity. These studies associated flipped instruction with improved conceptual understanding, higher-order thinking, and mathematics achievement. Their results suggest that flipped learning is most productive when students use classroom time to compare representations, explain relationships, justify procedures, and solve unfamiliar problems.

The current findings also clarify the role of self-regulation. Sun and Xie (2020) showed that learners exhibit different behavioral profiles during pre-class preparation, while Thai et al. (2017) demonstrated that learning outcomes depend on the combination of instructional explanation, guiding questions, and feedback. These results help explain why some flipped interventions underperform: access to digital content does not guarantee that students possess the planning, monitoring, persistence, and help-seeking skills required to learn independently.

Accordingly, the answer to RQ1 is affirmative but conditional. Flipped classroom interventions produce a moderate average improvement in mathematics learning, yet their effectiveness depends on whether the redistribution of instructional time is accompanied by substantive pedagogical transformation, adequate learner support, valid assessment, and coherent implementation.

5.2. Interpretation of RQ2: Evidence-Based Indicators of Quality

The second research question examined which indicators should be used to design, monitor, and evaluate flipped classroom teacher development in mathematics. The thematic synthesis identified seven mutually reinforcing dimensions: intentional pre-class design, mathematically rich in-class activity, formative assessment and feedback, integrated teacher knowledge, self-regulation support, equitable digital access, and analytics-informed adaptation.

These findings show that quality is systemic rather than additive. The indicators do not function as independent characteristics that can be implemented selectively without affecting the rest of the design. Instead, they form a causal and instructional chain. Pre-class preparation generates initial exposure and evidence of understanding; classroom tasks transform that preparation into mathematical activity; formative assessment reveals emerging difficulties; teacher knowledge supports interpretation and response; self-regulatory scaffolds enable participation; equitable access determines who can benefit; and learning analytics assists pedagogical adaptation.

A weakness in any component may reduce the effect of the entire intervention. High-quality videos cannot compensate for mathematically weak classroom tasks. Sophisticated dashboards cannot compensate for poor assessment design. Cognitively demanding activities cannot benefit learners who were unable to access or understand the preparatory material. Similarly, technological availability cannot compensate for insufficient pedagogical content knowledge or limited capacity to interpret student reasoning.

The findings extend the conventional FLIP framework by specifying the professional capacities required for mathematics education. A flexible environment, learning culture, intentional content, and professional educator remain relevant, but the present synthesis operationalizes these principles through observable and actionable indicators. In particular, the framework makes explicit the importance of mathematical task quality, formative use of evidence, self-regulation scaffolding, ethical analytics, and ecological feasibility.

The proposed indicators are also consistent with pedagogical content knowledge and mathematical knowledge for teaching. Shulman (1986, 1987) emphasized teachers' capacity to transform disciplinary knowledge into forms that learners can understand, while Ball et al. (2008) identified the specialized mathematical knowledge required to interpret strategies, diagnose errors, and select representations. In flipped learning, these capacities are required in both phases of instruction. Teachers must decide what content can be introduced independently and how it should be represented, while also designing classroom tasks capable of eliciting and extending student reasoning.

The framework further contributes by integrating TPACK with didactical suitability. Technology is treated as pedagogically valuable only when it improves access, representation, interaction, feedback, or evidence-informed adaptation. This interpretation avoids technologically deterministic accounts of innovation. The key question is not whether digital resources are present, but whether their use is mathematically meaningful, cognitively appropriate, interactionally productive, affectively supportive, and ecologically sustainable.

The integration of learning analytics constitutes another important contribution. Previous research has shown that analytics can reveal patterns of preparation, engagement, and participation in flipped environments (Algayres & Triantafyllou, 2019; Escudero-Nahón & Mercado-López, 2019). However, dashboards do not automatically improve instruction. Their value depends on teachers' capacity to formulate pedagogical questions, interpret data in context, connect traces with mathematical evidence, and act on the resulting interpretation.

The current framework therefore conceptualizes analytics as part of reflective instructional practice. The process begins with a pedagogical question, such as whether students understand a particular representation or concept. Relevant evidence is then examined, interpreted alongside student work and classroom interaction, and used to adapt grouping, pacing, feedback, or task selection. The effects of that adjustment subsequently become new evidence for reflection. In this sense, learning analytics is integrated into a cycle of professional inquiry rather than treated as a monitoring technology.

5.3. Contribution to Mathematics Teacher Professional Development

The findings support professional-development models organized around iterative cycles of design, enactment, evidence collection, collaborative analysis, and redesign. Such models are more compatible with flipped mathematics education than one-time technology workshops because they allow teachers to examine the relationship between digital preparation, classroom interaction, student reasoning, and learning outcomes.

Lesson Study is particularly well aligned with this process. Teachers can collaboratively identify mathematical learning problem, design pre-class and in-class tasks, anticipate student responses, observe implementation, analyses evidence, and revise the lesson. Video-based reflection and instructional coaching can further support the examination of teacher questions, explanations, feedback, and orchestration of discourse.

This conclusion is consistent with the research conducted by Morales-Maure et al. (2025c), which showed that Lesson Study and didactical suitability criteria can strengthen reflective judgement in mathematics teacher education. That study demonstrated that structured collaborative analysis can help teachers assess the epistemic, cognitive, interactional, mediational, affective, and ecological quality of instructional practice. The present meta-analysis extends that contribution by showing how these dimensions can be integrated with flipped learning and learning analytics.

The findings also converge with Morales-Maure et al. (2025b), who reported improvements in primary teachers' didactical-mathematical knowledge through task redesign in Panama. Both studies emphasize that professional learning is strengthened when teachers work directly with instructional tasks, analyses student responses, and redesign practice through theoretically informed criteria. The current study provides a broader international synthesis that situates these Panamanian findings within evidence from primary, secondary, university, and pre-service teacher education contexts.

This alignment demonstrates the international relevance of the research programme without implying national superiority. The Panamanian studies address the same core problems identified in international scholarship: how teachers transform mathematical content, how active-learning reforms are implemented with fidelity, how classroom evidence informs instructional decisions, and how professional development can produce sustainable pedagogical change. The contribution lies in showing that a research programme led from Panama can generate methodologically structured evidence that converges with international meta-analytic findings while introducing a distinctive integration of didactical suitability, Lesson Study, task redesign, teacher professional knowledge, and learning analytics.

The contribution of the research is therefore bidirectional. International meta-analytic evidence provides a comparative foundation for interpreting the Panamanian experience, while the Panamanian work contributes an integrated perspective based on didactical suitability, Lesson Study, task redesign, and teacher professional knowledge. This combination gives the research relevance beyond a single national context and offers a framework that may be adapted in Latin America and other education systems characterized by territorial, technological, and institutional diversity.

The international dimension is also reflected in the design of the FID2024-108 project, which articulates national evidence with international validation, collaboration with researchers from the University of Barcelona, and comparative examination of teacher professional development. The project has generated peer-reviewed outputs concerning Lesson Study, didactical suitability, and task redesign, thereby connecting locally developed evidence with international debates in mathematics education.

Nevertheless, international impact should not be equated solely with geographical dissemination or publication. Its stronger scientific expression lies in whether the proposed framework contributes concepts, methods, and indicators that are transferable, testable, and useful across educational contexts. The seven-indicator framework advances this objective by translating the heterogeneous flipped-learning literature into a structured model for programme design, observation, evaluation, and policy analysis.

5.4. Theoretical Contribution

The study contributes theoretically by integrating domains that are often treated separately. Flipped classroom research frequently focuses on instructional sequencing or technology use; teacher-development research focuses on professional knowledge and collaborative learning; mathematics education emphasizes task quality, reasoning, and didactical suitability; and learning analytics examines digital traces and data-informed decision-making.

The present framework connects these perspectives within a single explanatory model. It proposes that flipped learning effectiveness emerges from coherence among mathematical quality, pedagogical alignment, professional knowledge, self-regulation, interaction, analytics, equity, and institutional sustainability.

This integration helps explain why interventions sharing the same label generate markedly different results. “Flipped classroom” is not a homogeneous treatment but a family of designs whose pedagogical mechanisms vary considerably. The study therefore shifts the unit of interpretation from the presence of a flipped format to the quality of the complete didactical system.

The framework also strengthens the construct of didactical suitability by incorporating digital evidence and analytics. Epistemic and cognitive suitability determine which mathematical processes should be monitored; interactional suitability informs the interpretation of discourse and participation; mediational suitability addresses platforms and access; affective suitability directs attention to motivation and agency; and ecological suitability situates analytics within institutional, ethical, and social constraints.

5.5. International and Policy Relevance

The results have direct relevance for teacher-education institutions, universities, ministries of education, and professional-development providers. The moderate average effect indicates that flipped learning represents a defensible instructional option, but the high heterogeneity warns against large-scale adoption based solely on technology acquisition or standardised content production.

Policies should therefore prioritise teacher capacity rather than platform availability. Investment in recording equipment, learning-management systems, or dashboards will have limited educational value unless accompanied by sustained development in mathematical task design, formative assessment, instructional orchestration, data interpretation, and inclusive design.

For countries with unequal connectivity and geographically dispersed populations, including Panama and other Latin American contexts, the equity indicator is especially important. Flipped models may increase flexibility, but they may also reproduce or deepen inequality when participation depends on devices, stable internet access, time, and family support. Low-bandwidth materials, downloadable resources, flexible access, captions, and alternative participation routes should therefore be treated as minimum quality standards.

The framework also offers a basis for cross-national comparison. Rather than comparing programmes only through mean achievement gains, future evaluations can examine how implementation differs across the seven dimensions. This would permit more meaningful comparison among countries, educational levels, and institutional contexts.

5.5.1. Legal and Institutional Pathways for Turning Quality Indicators into University Educational Public Policy in Panama

Converting the seven indicators into enforceable public policy requires attention to the legal architecture governing Panamanian higher education. University autonomy has constitutional rank: Article 103 of the Constitución Política de la República de Panamá establishes that "la Universidad Oficial de la República es autónoma," a guarantee developed by the Ley Orgánica de la Universidad de Panamá (Ley 24 de 14 de julio de 2005), whose Articles 3 and 48 recognise the university's power of self-regulation (autorreglamentación) over academic and curricular matters. This means that adoption of the seven indicators in teacher-education programmes cannot be imposed top-down; it must proceed through the university's own governance bodies, consistent with the deliberative model the institution itself invokes to defend its constitutional autonomy.

At the national level, the indicators intersect with an existing quality-assurance instrument. The Sistema Nacional de Evaluación y Acreditación para el Mejoramiento de la Calidad de la Educación Superior Universitaria de Panamá, created by Ley 52 de 26 de junio de 2015 and regulated by Decreto Ejecutivo N.º 539 de 30 de agosto de 2018, empowers CONEAUPA the council that governs and administers that system to define the methodological criteria used in institutional and programme accreditation, which is compulsory for official universities. Embedding the seven evidence-based indicators into CONEAUPA's evaluation instruments for teacher-education and mathematics-education programmes would give the framework binding, auditable status, rather than leaving it as an academic recommendation with no regulatory channel.

A second channel runs through teacher-training governance. The Ley 47 de 1946 (Ley Orgánica de Educación) assigns the Ministerio de Educación (MEDUCA), jointly with the official universities, responsibility for coordinating and organising initial and continuing teacher training. Because the University of Panama is the historical trainer of the national teaching corps, this shared competence provides a concrete legal pathway for proposing the seven indicators as minimum quality standards for MEDUCA-recognised professional-development programmes in mathematics education, extending their reach beyond individual institutional practice into national teacher-policy design.

Finally, the equity indicator identified in the meta-analysis acquires constitutional weight in this framework. Article 91 of the Constitución Política establishes education as a right of every person and a public service organised by the State; unequal access to connectivity and devices among flipped-classroom participants therefore raises not only a pedagogical concern but a constitutional one. Framing equitable access as a normative obligation, rather than only good practice, strengthens the case for conditioning public and institutional investment in flipped-classroom teacher development on demonstrated compliance with the equity indicator an accountability logic consistent with SENACYT's own funding standards under Project FID2024-108.

6. Conclusions

This systematic review and meta-analysis provides evidence that flipped classroom interventions have a moderate positive effect on mathematics learning. The pooled estimate from 19 studies with recoverable uncertainty information was $d = 0.522$, and the independent synthesis reported by Yakar (2021) produced a closely aligned estimate of $d = 0.510$. The consistency of these results with prior meta-analyses by Güler et al. (2023), Strelan et al. (2020), and Haziki et al. (2025) supports the conclusion that flipped learning can improve mathematics achievement across school, university, and teacher-education contexts.

However, the very high heterogeneity observed in both quantitative syntheses demonstrates that effectiveness is not inherent in the flipped format. The educational value of the model depends on the quality and coherence of its implementation. Flipped learning produces stronger outcomes when pre-class resources are concise and aligned with mathematical goals, classroom time is devoted to demanding mathematical activity, formative evidence guides instruction, teachers possess integrated mathematical-pedagogical-technological knowledge, self-regulation is explicitly scaffolded, access barriers are addressed, and analytics are used ethically to support adaptation.

The study therefore rejects a technologically reductionist interpretation of flipped learning. Videos, platforms, and dashboards are not sufficient indicators of pedagogical innovation. The defining criterion is whether the instructional system creates better opportunities for learners to reason, communicate, represent, solve problems, receive feedback, and progressively regulate their own learning.

The seven evidence-based quality indicators developed in this study constitute its principal theoretical and practical contribution. They provide a framework for designing professional-development programmes, evaluating flipped mathematics courses, observing teaching practice, assessing teacher portfolios, and informing institutional or public-policy decisions. They also offer an explanation for the substantial between-study variability that characterizes international evidence. In Panama, this translates concretely into three levers: incorporation of the seven indicators into CONEAUPA's accreditation criteria for teacher-education programmes (Ley 52 de 2015), articulation with MEDUCA's teacher-training mandate under the Ley Orgánica de Educación (Ley 47 de 1946), and conditioning of public research and development funding, such as SENACYT's Project FID2024-108, on demonstrated compliance with the equity indicator.

The study further demonstrates the international relevance of a research programme led from Panama. Its findings converge with international evidence while contributing a distinctive integration of flipped learning, mathematics teacher knowledge, Lesson Study, didactical suitability, task redesign, and learning analytics. The prior studies by Morales-Maure et al. (2025b, 2025c) provide empirical evidence that collaborative reflection and task redesign can strengthen mathematics teachers' professional knowledge. The present synthesis extends that work by positioning it within a comparative international evidence base and translating it into a transferable framework of quality indicators.

Accordingly, the scientific contribution is not limited to validating a pedagogical strategy in one national context. It provides an analytical model with potential application across diverse education systems, particularly those seeking to combine active learning, digital transformation, teacher professional development, and equity. This international applicability should be examined through future cross-country studies, longitudinal designs, and implementation research in which the seven indicators are operationalized and tested as moderators of teacher and student outcomes. Flipped classroom instruction should be understood as a professionally demanding model of pedagogical reorganization rather than as a simple inversion of teaching time. Its success depends on teachers' capacity to transform content, interpret evidence, orchestrate mathematical interaction, and adapt instruction to learner and contextual needs. When these conditions are present, flipped learning can contribute meaningfully to mathematics achievement, teacher learning, and sustainable educational innovation.

Academic Contribution. This study contributes to mathematics teacher-education research by demonstrating that the effectiveness of flipped learning depends less on the relocation of content than on the coherence among mathematical task quality, formative assessment, teacher professional knowledge, learner self-regulation, digital equity, and analytics-informed adaptation. The study extends existing flipped-classroom and teacher-development frameworks by integrating seven evidence-based quality indicators into a transferable model for mathematics education, while methodologically combining meta-analysis, comparative validation, and thematic synthesis to explain substantial between-study heterogeneity. Empirically, it adds evidence generated from a Panamanian-led research programme to the international literature on mathematics teacher professional development and higher education practice.

Author Contributions: Conceptualization, L.M.-M., E.G.V., O.G.-M., and M.Á.E.; methodology, L.M.-M., E.G.V., and O.G.-M.; formal analysis, L.M.-M., E.G.V., and M.Á.E.; investigation, L.M.-M., E.G.V., O.G.-M., and M.Á.E.; data curation, L.M.-M. and E.G.V.; writing—original draft preparation, L.M.-M.; writing—review and editing, E.G.V., O.G.-M., and M.Á.E.; visualization, L.M.-M. and O.G.-M.; supervision, E.G.V. and M.Á.E.; project administration, L.M.-M. and E.G.V.; funding acquisition, L.M.-M., E.G.V., O.G.-M., and M.Á.E. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest. No financial or personal relationship influenced the design, analysis, interpretation, or reporting of the study.

Funding: This study was supported by the Secretaría Nacional de Ciencia, Tecnología e Innovación of Panama (SENACYT) through Project FID2024-108, Contract No. 096-2024, and by institutional support from the Universidad de Panamá. The funders had no involvement in the study design, literature screening, data extraction, meta-analytic procedures, interpretation of results, manuscript preparation, or publication decision.

Acknowledgments: The authors gratefully acknowledge the Universidad de Panamá and SENACYT–Panama for their institutional and scientific support. The authors also thank the national and international collaborators who provided methodological guidance, statistical review, and academic discussion throughout the development of Project FID2024-108, including colleagues associated with the mathematics-education research network and the University of Barcelona. Appreciation is also extended to the research team members who supported database searching, screening, data extraction, and verification of the meta-analytic evidence.

Bibliography

1. Algayres, M., & Triantafyllou, E. (2020). Learning analytics in flipped classrooms: A scoping review. Electronic Journal of e-Learning, 17(2), 99–112. https://doi.org/10.34190/JEL.18.5.003
2. Asamblea Nacional de Panamá. (1946). Ley 47, Orgánica de Educación (texto único, Decreto Ejecutivo N.° 305 de 2004). https://vlex.com.pa/vid/ley-47-1946-ley-1049312809
3. Asamblea Nacional de Panamá. (2005). Ley 24 de 14 de julio de 2005, Orgánica de la Universidad de Panamá. Gaceta Oficial N.° 25,344. https://www.up.ac.pa/sites/default/files/2021-07/LeyOrganica.pdf
4. Asamblea Nacional de Panamá. (2015). Ley 52 de 26 de junio de 2015, que crea el Sistema Nacional de Evaluación y Acreditación para el Mejoramiento de la Calidad de la Educación Superior Universitaria de Panamá. https://coneaupa.edu.pa/acuerdos-leyes/
5. Atta, S. A., & Bonyah, E. (2023). Designing a flipped classroom instruction to improve plane geometry learning among pre-service teachers in Ghana. Contemporary Mathematics and Science Education, 4(1), ep23004. https://doi.org/10.30935/conmaths/12674
6. Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? Journal of Teacher Education, 59(5), 389–407. https://doi.org/10.1177/0022487108324554
7. Bhagat, K. K., Chang, C. N., & Chang, C. Y. (2016). The impact of the flipped classroom on mathematics concept learning in high school. Educational Technology & Society, 19(3), 134–142. https://doi.org/10.30191/ETS.201607_19(3).0013
8. Blau, I., & Shamir-Inbal, T. (2017). Re-designed flipped learning model in an academic course: The role of co-creation and co-regulation. Computers & Education, 115, 69–81. https://doi.org/10.1016/j.compedu.2017.07.014
9. 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
10. CONEAUPA. Decreto Ejecutivo N.° 539 de 30 de agosto de 2018, que reglamenta la Ley 52 de 2015. https://coneaupa.edu.pa/instrumentos-legales/

11. Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development. Palo Alto, CA: Learning Policy Institute. https://eric.ed.gov/?id=ED606743
12. Escudero-Nahón, A., & Mercado-López, E. P. (2019). Use of learning analytics in the flipped classroom: A systematic review. Apertura, 11(2), 72–85. https://doi.org/10.32870/Ap.v11n2.1546 https://doi.org/10.32870/Ap.v11n2.1546
13. Fatmiyati, N., Juandi, D., & Fatimah, S. (2026). Meta-analysis of flipped classroom model to promote mathematical higher-order thinking. Journal of Education and Learning (EduLearn), 20(2), 945-954.
14. Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5–6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816
15. Güler, M., Kokoç, M., & Önder Bütüner, S. (2023). Does a flipped classroom model work in mathematics education? A meta-analysis. Education and Information Technologies, 28(1), 57-79. https://doi.org/10.1007/s10639-022-11143-z
16. Haziki, N. H., Abdullah, A. H., & Hamzah, M. H. (2025). Flipped Classrooms in Higher Education: A Meta-Analysis of Their Impact on Mathematics Performance. Malaysian Journal of Mathematical Sciences, 19(1), 125–146. https://doi.org/10.47836/mjms.19.1.07
17. Helsa, Y., Turmudi, & Juandi, D. (2023). Hybrid learning for conceptual understanding skills in mathematics: A meta-analysis. International Journal of Information and Education Technology, 13(2), 355–363. https://doi.org/10.18178/ijiet.2023.13.2.1814
18. Hwang, G.-J., Lai, C.-L., & Wang, S.-Y. (2015). Seamless flipped learning: A mobile technology-enhanced flipped classroom with effective learning strategies. Journal of Computers in Education, 2, 449–473. https://doi.org/10.1007/s40692-015-0043-0
19. Karjanto, N., & Acelajado, M. J. (2022). Sustainable learning, cognitive gains, and improved attitudes in college algebra flipped classrooms. Sustainability, 14(19), 12500. https://doi.org/10.3390/su141912500
20. Kaya, D. (2018). The effect of flipped learning model on middle school students' classroom engagement in teaching mathematics. Sakarya University Journal of Education, 8(4), 232–249. https://doi.org/10.19126/suje.453729
21. Love, B., Hodge, A., Grandgenett, N., & Swift, A. W. (2014). Student learning and perceptions in a flipped linear algebra course. International Journal of Mathematical Education in Science and Technology, 45(3), 317–324. https://doi.org/10.1080/0020739X.2013.822582
22. Ministerio Público de Panamá. Constitución Política de la República de Panamá con índice analítico. https://ministeriopublico.gob.pa/wp-content/uploads/2016/09/constitucion-politica-con-indice-analitico.pdf
23. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record: The Voice of Scholarship in Education, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x
24. Moon, H., Seo, Y. N., & Kang, S. A. (2026). Rethinking training governance in South Korea: a meta-analytical approach to innovative supply models. European Journal of Training and Development, 1-18.
25. Morales-Maure, L., Chacón-Rivadeneira, K., Garcia-Marimón, O., Sáez-Delgado, F., & Campos-Nava, M. (2025a). Enhancing Mathematics Teacher Training in Higher Education: The Role of Lesson Study and Didactic Suitability Criteria in Pedagogical Innovation. Trends in Higher Education, 4(3), 39. https://doi.org/10.3390/higheredu4030039
26. Morales-Maure, L., Torres-Rodríguez, A. A., Campos-Nava, M., & Martínez, I. V. (2025b). Fortalecimiento del conocimiento didáctico matemático en maestros de primaria en Panamá a través del rediseño de tareas. Formación Universitaria, 18(1), 89–100.
27. Morales-Maure, L., Vazquez, E. G., García Marimón, O., & Alfaro Ponce, B. (2025c). Assessment of the impact of the Lesson Study approach and didactic suitability criteria on mathematics teacher training from the onto-semiotic approach. Edelweiss Applied Science and Technology, 9(4), 602–616.
28. Nurhajarurahmah, S. Z., Syarifuddin, S., & Ismiyati, N. (2026). The effect of flipped learning models on students' mathematical abilities: A meta-analysis study. Jurnal Pendidikan dan Pembelajaran Indonesia (JPPI), 6(2), 1849–1864. https://doi.org/10.53299/jppi.v6i2.4441
29. Ruiz-Palmero, J., Guillén-Gámez, F. D., Colomo-Magaña, E., & Sánchez-Vega, E. (2023). Effectiveness of the flipped classroom in the teaching of mathematics in an online environment: Identification of

factors affecting the learning process. Online Learning, 27(2), 304–326. https://eric.ed.gov/?id=EJ1393112

30. Sacristán San Cristóbal, M., Martín R, D., Navarro Asencio, E. & Tourón, J. (2017). Flipped classroom y didáctica de las matemáticas en la formación online de Maestros de Educación Infantil. Revista Electrónica Interuniversitaria de Formación del Profesorado, 20(3), 1-14. https://reunir.unir.net/handle/123456789/6375
31. Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4–14. https://doi.org/10.3102/0013189X015002004
32. Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review, 57(1), 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411
33. Siemens, G., & Baker, R. S. D. (2012). Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254). https://doi.org/10.1145/2330601.2330661
34. Song, Y., & Kapur, M. (2017). How to flip the classroom: Productive failure or traditional flipped classroom pedagogical design? Educational Technology & Society, 20(1), 292–305. https://www.jstor.org/stable/jeductechsoci.20.1.292
35. 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
36. Sun, Z., & Xie, K. (2020). How do students prepare in the pre-class setting of a flipped undergraduate math course? A latent profile analysis of learning behavior and the impact of achievement goals. The Internet and Higher Education, 46, 100731. https://doi.org/10.1016/j.iheduc.2020.100731
37. Surjono, H. D., & Mahmudi, A. (2023). The effect of flipped classroom learning model on students' understanding of mathematical concepts and higher-order thinking skills. International Journal of Information and Education Technology, 13(12), 2014–2022. https://doi.org/10.18178/ijiet.2023.13.12.2016
38. Thai, N. T. T., De Wever, B., & Valcke, M. (2017). The impact of a flipped classroom design on learning performance in higher education: Looking for the best blend of lectures and guiding questions with feedback. Computers & Education, 107, 113–126. https://doi.org/10.1016/j.compedu.2017.01.003
39. Toivola, M., Rajala, A., & Kumpulainen, K. (2023). Pedagogical rationales of flipped learning in the accounts of Finnish mathematics teachers. Pedagogies: An International Journal, 18(4), 767–787. https://doi.org/10.1080/1554480X.2022.2077341
40. Tribunal Electoral de Panamá. Constitución Política de la República de Panamá, vigente. https://constitucion.te.gob.pa/constitucion-vigente/
41. van Leeuwen, A. (2019). Teachers' perceptions of the usability of learning analytics reports in a flipped university course: When and how does information become actionable knowledge? Educational Technology Research and Development, 67(3), 701–720. https://doi.org/10.1007/s11423-018-09639-y
42. Wasserman, N. H., Quint, C., Norris, S. A., & Carr, T. (2017). Exploring flipped classroom instruction in Calculus III. International Journal of Science and Mathematics Education, 15, 545–568. https://doi.org/10.1007/s10763-015-9704-8
43. Webel, C., Sheffel, C., & Conner, K. A. (2018). Flipping instruction in a fifth grade class: A case of an elementary mathematics specialist. Teaching and Teacher Education, 71, 271–282. https://doi.org/10.1016/j.tate.2018.01.007
44. Wei, X., Cheng, I. L., Chen, N. S., Yang, X., Liu, Y., Dong, Y., ... & Kinshuk. (2020). Effect of the flipped classroom on the mathematics performance of middle school students. Educational Technology Research and Development, 68(3), 1461-1484. https://doi.org/10.1007/s11423-020-09752-x
45. Yakar, Z. Y. (2021). The effect of flipped learning model on primary and secondary school students' mathematics achievement: A meta-analysis study. Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 50(2), 1329-1366. https://doi.org/10.14812/cuefd.865337
46. Yang, Q., & Chen, C. (2026). The effectiveness of flipped classroom on autonomous learning ability: a meta-analysis. Asia Pacific Education Review, 27(2), 705.
47. 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

Luisa Morales-Maure holds a PhD in Mathematics and Experimental Sciences Education from the University of Barcelona. She is a professor and researcher at the University of Panama and coordinator of the Mathematics Education Research Group GIEM-21. Her research focuses on mathematics education, teacher professional development, didactic suitability, STEM education, digital technologies, and artificial intelligence in education. She is a Level I member of Panama's National Research System.

Evelyn García Vázquez holds a doctorate in Education and is a visiting professor and researcher at the University of Panama. Her academic work focuses on higher education, organizational culture, inclusive education, educational assessment, and mathematics education. She has participated in interdisciplinary and international research projects examining teacher development, reflective practice, educational innovation, and the improvement of mathematics teaching and learning.

Keila Chacón-Rivadeneira is a doctoral researcher in Mathematics Education at the University of Barcelona. She holds postgraduate qualifications in Mathematics Education, Educational Technology, and Digital Competencies. Her research focuses on mathematics teacher education, didactic suitability, digital competence, and the pedagogical integration of artificial intelligence.

Berenice Alfaro-Ponce holds a PhD in Social Sciences from the Autonomous University of the State of Hidalgo and is an associate researcher at Tecnológico de Monterrey. Her research addresses education and technology policy, human rights, gender equality, social inclusion, and sustainable development. She is a Level I member of Mexico's National System of Researchers and has contributed to interdisciplinary projects connecting educational innovation, public policy, and social transformation.

Orlando García-Marimón holds a PhD in Didactics of Sciences, Arts, and Humanities from the University of Barcelona. He is an associate professor and researcher at the University of Panama and a member of the GIEM-21 research group. His research interests include mathematics and science education, teacher professional development, reflective practice, hybrid learning, and technology-enhanced education. He is a Level I member of Panama's National Research System.

Miguel Ángel Esbrí Montoliu holds a PhD in Law from the Universitat Jaume I (Spain) and two master's degrees in finance, from the Universidad Internacional de Cataluña and the Universidad Francisco de Vitoria, Spain. He is a professor of law in the Department of Social Law, Faculty of Law and Political Sciences, at the Universidad de Panamá, and a Category II member of Panama's National Research System (SNI), accredited by SENACYT. His academic work focuses on law and educational policy, including the legal and institutional frameworks that govern higher-education quality assurance in Panama. He is member of the GIEM-21 research group.