Available online at: http://journal.uny.ac.id/index.php/jpms

Jurnal Pendidikan Matematika dan Sains, 14 (2), 2026, 357–371

A Comparative Meta-analysis of the Effects of Gamification and Flipped Classroom on Secondary School Students' Mathematics Achievement

Amelia Vega *, Erry Hidayanto, Makbul Muksar
Mathematics Education Department, Universitas Negeri Malang, Indonesia
* Corresponding Author. E-mail: amelia.vega.2203118@students.um.ac.id

Abstract

This study compares the effectiveness of gamification and the flipped classroom on secondary school students' mathematics achievement through a systematic review and meta-analysis. The synthesis follows PRISMA 2020 and includes 16 quasi-experimental studies published between 2015 and 2025, comprising 5 gamification studies and 11 flipped classroom studies. Effect sizes were calculated using Hedges' g and pooled using a random-effects model with the Hartung-Knapp-Sidik-Jonkman (HKSJ) adjustment to accommodate heterogeneity. The results show that gamification has a small-to-moderate positive point estimate ($g = 0.437$), but the confidence interval crosses zero under high heterogeneity ($I^2 \approx 87\%$), indicating that the evidence for benefit is uncertain. The flipped classroom shows a moderate-to-large pooled effect ($g = 1.009$) with a significant confidence interval, although heterogeneity is also high ($I^2 \approx 86\%$). Subgroup analysis found a significant difference between the two approaches, but meta-regression indicates that model type is not a significant predictor after between-study variation is accounted for. Overall, the findings suggest that effectiveness is design-sensitive and influenced more by implementation quality and context than by the instructional label alone.

Keywords: *Flipped classroom, Gamification, Meta-analysis, Mathematics education*

How to Cite: Vega, A., Hidayanto, E., & Muksar, M. (2026). A comparative meta-analysis of the effects of gamification and flipped classroom on secondary school students' mathematics achievement. *Jurnal Pendidikan Matematika dan Sains, 14*(2). 357–371. https://doi.org/10.21831/jpms.v14i2.96112

DOI: https://doi.org/10.21831/jpms.v14i2.96112

INTRODUCTION

Mathematics learning at the secondary school level continues to face persistent challenges, as many students struggle to make sense of abstract forms, retain conceptual understanding, and manage mathematics anxiety, all of which are associated with lower achievement and less productive learning experiences (Caviola et al., 2022; Demedts et al., 2022). This condition is also reflected in studies published in the *Jurnal Pendidikan Matematika dan Sains (JPMS)*, which report that students' difficulties in mathematics are closely related to conceptual understanding and mathematics anxiety, both of which influence learning outcomes (Asanre et al., 2025; Sitaresmi & Hanifah, 2026). In addition, mathematics instruction in classrooms is still often criticized for placing too much emphasis on teacher explanation and mechanical solution procedures, even though better learning outcomes are more likely to be achieved when students actively participate in meaning-making, discussion, and problem solving (Zajda, 2023). These characteristics make secondary school mathematics a highly important context for examining instructional designs that not only improve achievement, but also enhance the quality of students' cognitive and motivational engagement (Demedts et al., 2022). Among technology-assisted instructional innovations, gamification and the flipped classroom have received sustained attention because both seek to improve learning through different pedagogical mechanisms (L. Li et al., 2024; Ma et al., 2025).

Gamification integrates game elements such as points, badges, leaderboards, challenges, and feedback into learning activities in order to encourage participation and persistence. However, its contribution to learning is determined not merely by the presence of game elements, but by whether the design supports students' autonomy, competence, and social relatedness (M. Li et al., 2023; Wang et al., 2024). This argument is consistent with Self-Determination Theory, which predicts that motivational support is more likely to foster meaningful engagement when students are given opportunities to make choices, experience competence, and build social connections during learning (Wang et al., 2024). Recent evidence also indicates that the effects of gamification are mixed. Although meta-analyses generally report positive effects on learning, the magnitude of these benefits varies substantially across studies and may weaken when competition, rewards, or performance pressure are more prominent than meaningful learning processes (L. Li et al., 2024).

In contrast, the flipped classroom restructures when and how students engage with mathematical content by shifting initial exposure to foundational material to the pre-class phase and using face-to-face class time for guided practice, discussion, feedback, and problem solving (Ma et al., 2025; Tutal & Yazar, 2021). This structure is theoretically aligned with constructivist, active learning, and social constructivist perspectives because it creates more opportunities for students to build understanding through interaction, explanation, and supported participation during classroom learning (Zajda, 2023). This approach is also consistent with Cognitive Load Theory, as pre-class materials can help students process foundational information at their own pace, while in-class activities can focus on deeper conceptual understanding and higher-order mathematical thinking (Fischer et al., 2023). Recent meta-analytic evidence confirms that the flipped classroom tends to improve learning outcomes, including in mathematics, but it also reveals substantial variation depending on the quality of pre-class and in-class activity design, disciplinary context, and the type of outcomes measured (Güler et al., 2023; Tutal & Yazar, 2021).

Although this body of research continues to grow, the existing evidence base remains insufficient to answer a more specific question that is theoretically and practically important: how do gamification and the flipped classroom compare when the focus is limited to secondary school mathematics learning (Güler et al., 2023; L. Li et al., 2024; Ma et al., 2025). Existing meta-analyses have mostly examined each approach separately, often across broad disciplinary or educational contexts, and therefore do not yet sufficiently clarify whether differences in effectiveness reflect the instructional models themselves or the quality of the instructional design embedded within them (L. Li et al., 2024; Ma et al., 2025). This limitation is particularly important in mathematics learning, because learning in this field depends on the gradual development of conceptual understanding, representational fluency, and problem-solving strategies, all of which may be influenced differently by motivational game elements and by the reallocation of instructional time in flipped learning (Demedts et al., 2022; Güler et al., 2023). Therefore, a comparative synthesis in this domain has value not only for instructional decision-making, but also for theory development, because it allows the respective roles of motivational and cognitive-active mechanisms to be examined within the same field of study (L. Li et al., 2024; Ma et al., 2025). Accordingly, this study conducts a systematic literature review with quantitative synthesis to compare the effectiveness of gamification and the flipped classroom in secondary school mathematics learning, following PRISMA 2020 procedures and using a random-effects meta-analytic model (Higgins et al., 2019; Page et al., 2021). In addition to estimating pooled effects, this study examines whether variation in findings is related to instructional design and implementation context, thereby clarifying that the effectiveness of these approaches should be interpreted not merely as a function of the model label, but as a function of how the learning experience is designed (L. Li et al., 2024; Ma et al., 2025). Thus, this study provides a subject-specific comparative synthesis for secondary school mathematics, strengthens the theoretical framework of two widely used instructional approaches, and offers evidence to support more context-sensitive and design-based pedagogical decision-making (Güler et al., 2023; L. Li et al., 2024; Ma et al., 2025). To the best of our knowledge, there has been no meta-analysis that directly compares gamification and the flipped classroom within secondary mathematics education using a single comparative synthesis.

METHODS

This study employed a systematic literature review (SLR) with meta-analysis to compare the effectiveness of gamification and the flipped classroom on secondary school students' mathematics achievement. The review was reported in accordance with the PRISMA 2020 guidelines to ensure transparency in the processes of study identification, screening, eligibility assessment, and reporting (Page et al., 2021). External protocol registration was not undertaken for this study due to time constraints and practical considerations surrounding the registration process. However, prior to the screening stage, the review procedures had already been established, including the research focus, inclusion and exclusion criteria, search strategy, screening process, data extraction instrument, instrument validation, methodological quality assessment, and statistical analysis plan. Therefore, although the review was not externally registered, it still followed predefined methodological procedures in line with the PRISMA 2020 framework to maintain transparency and minimize post-hoc analytical decisions. Quantitative synthesis was conducted through meta-analysis. Because the effects across studies were assumed to vary due to differences in educational context, participant characteristics, and intervention design, the pooled effect size was calculated using a random-effects model (Higgins et al., 2019).

The literature search was conducted across 11 databases: SpringerLink, ScienceDirect, Wiley, ProQuest, SAGE Journals, Taylor & Francis, Cambridge Journals, Emerald Insight, Nature Portfolio, JSTOR, and Oxford University Press. Searches across the selected databases were conducted during September 2025, and the final search update was completed on 2 October 2025. Across all databases, the search was limited to studies published between 1 January 2015 and 2 October 2025, in line with the predefined eligibility criteria. Accordingly, only studies published up to that date were considered for inclusion in this review. The following search string was used across all databases: ("gamification" OR "flipped classroom") AND "quasi-experimental" AND ("secondary school" OR "high school") AND (mathematics OR "mathematics education"). The search strategy was applied consistently across databases, with minor adjustments to syntax and filtering options based on the search features available in each database. The only differences concerned the adjustment of filters according to the features available in each database. The filters applied consistently included publication year (2015–2025), English language, and research article/full-text availability. By applying these filters, non-article sources and grey literature, such as theses, dissertations, conference proceedings, reports, and preprints, were excluded from the search results. No manual reference-list screening was performed. Therefore, the study identification process relied on database searching only, which may have increased the possibility that some relevant studies not captured by the indexed search results were missed. All retrieved records were exported from the databases in downloadable formats, including Excel and CSV files, and merged into a single dataset before screening. Deduplication was performed by first combining all search results and then standardizing the information, i.e., writing all text in the same format. Afterward, a Python program in Google Colab was used to find and remove duplicate records. The program examined key details such as title, author name, and year of publication. The number of duplicates found was relatively small because the search was initially limited to specific years, English-language journal articles, full-text papers, and topics related to this research. Furthermore, different databases often store article information in slightly different ways, so some similar records were not detected as exact duplicates by the program. Instead, such records were more often removed during the title, abstract, and full-text screening stages. After that, study selection was conducted in two stages: title and abstract screening followed by full-text screening. Because the screening and methodological appraisal were conducted by a single reviewer, additional consistency checks were applied to reduce subjectivity. The screening process was conducted in two separate rounds at different times, and full-text eligibility decisions were rechecked against the predefined inclusion and exclusion criteria.

The inclusion criteria were established to maintain consistency in study selection. Studies were included if they used an experimental or quasi-experimental design, applied gamification or the flipped classroom in mathematics learning, involved secondary school students, and reported quantitative data that allowed effect-size calculation. In addition, only English-language journal articles with full-text availability were included in the synthesis. The focus on experimental and quasi-experimental studies was chosen because this meta-analysis aimed to estimate intervention effects based on comparative quantitative data between groups. Studies that lacked a comparison group or did not provide sufficient statistical data were excluded because they did not permit standardized effect-size calculation. Although both experimental and quasi-experimental studies were conceptually eligible, all studies that ultimately met the inclusion criteria in this review used a quasi-experimental design. The selection process followed the four PRISMA stages: identification, screening, eligibility, and inclusion. Of the 776 articles identified at the initial stage, 16 studies met all criteria and were included in the meta-analysis.

Data from each study were collected using a standardized extraction sheet containing information on the authors, publication year, country, sample size, type of intervention, and statistical data from the experimental and control groups required for effect-size calculation. The data extraction instrument was first subjected to content validation through expert judgment. The validation was conducted by one expert validator using a 4-point scale, namely 1=not relevant, 2=less relevant, 3=fairly relevant, and 4=very relevant. The validation covered six aspects: clarity of the SLR title and research question, clarity of the inclusion and exclusion criteria, clarity of the article screening instrument, clarity of the article extraction instrument, appropriateness of terms and language, and ease of instrument use. The instrument obtained an average score of 3.5, which falls into the very valid category within the range of 3.25–4.00, indicating that it was feasible for use without revision. Effect sizes were calculated using the standardized mean difference (SMD) with Hedges' g to correct for bias in small samples (Higgins et al., 2019). The interpretation of effect sizes followed the conventional SMD thresholds of 0.20 (small), 0.50 (moderate), and 0.80 (large), while still considering the context of the studies analyzed (Andrade, 2023).

The methodological quality of the selected studies was assessed using the JBI Critical Appraisal Checklist for Quasi-Experimental Studies, which, according to Barker et al. (2024), is used to assess the risk of bias in quasi-experimental studies through nine appraisal items (Q1–Q9). Each item was rated using the categories Y (Yes), N (No), U (Unclear), or NA (Not applicable) in accordance with the JBI checklist user guide. The quality assessment was conducted by one reviewer. The results of the JBI appraisal were used to help explain the magnitude of the effect sizes and the level of variation in findings across studies included in the meta-analysis.

Statistical analysis was conducted in Python using Google Colab. The packages used in the analysis included pandas, numpy, matplotlib, seaborn, scipy, and statsmodels. The pooled effect size was calculated using a random-effects model with an estimate of between-study variance (τ^2). Because the number of included studies was relatively small, confidence intervals were also examined using the Hartung-Knapp-Sidik-Jonkman (HKSJ) approach to provide more robust interval estimates. Heterogeneity was assessed using Cochran's Q statistic and the I^2 index, with I^2 values of 25%, 50%, and 75% indicating low, moderate, and high heterogeneity, respectively (Hegazi et al., 2025). The value of τ^2 was also calculated and reported as an estimate of between-study variability. Publication bias was assessed visually through a funnel plot and statistically tested using Egger's regression test and Begg's rank correlation test, both for the overall analysis and for each subgroup. In addition, sensitivity analysis was conducted in two ways. First, the pooled effect estimate was compared before and after excluding studies identified as methodologically weaker based on the risk-of-bias assessment. Second, for the gamification subgroup, an additional sensitivity check was performed by excluding the study that appeared most influential and inconsistent with the overall pattern, in order to assess the stability of the pooled estimate and heterogeneity.

RESULTS AND DISCUSSION

The literature selection process followed the PRISMA 2020 guidelines to ensure a systematic review that is transparent and replicable (Page et al., 2021). The flow of identification, screening, and study inclusion is presented in Figure 1. of the 776 articles identified, 16 studies met all inclusion criteria and were included in the meta-analysis. Only 18 records were removed at the deduplication stage, which reflects the restrictive search filters and differences in metadata formatting across databases, while duplicate detection itself was conducted through bibliographic harmonization and automated checking in Python. All studies used a quasi-experimental design with a control group, which is commonly used when randomized controlled trials (RCTs) are not feasible in research on instructional strategies (Cham et al., 2024).

Although the broader literature on gamification and flipped classroom continues to expand, only a limited number of studies met the eligibility criteria of this review. This was mainly due to the narrow scope of the synthesis, which was restricted to secondary school mathematics, English-language journal articles published between 2015 and 2025, quasi-experimental or experimental comparison designs, and studies reporting sufficient quantitative data for effect-size calculation. As a result, many studies in the wider literature were excluded because they focused on other subjects, different educational levels, non-comparative designs, or incomplete statistical reporting. However, when only a small number of studies are included, the results need to be interpreted more cautiously because uncertainty and between-study differences may be greater (Schulz et al., 2022). A summary of the characteristics of the included studies is shown in Table 1 and 2, including intervention duration, the gamification elements used in gamified studies, and the types of pre-class materials used in flipped classroom studies. Of all studies, five implemented gamification and eleven implemented the flipped classroom in secondary school mathematics learning.

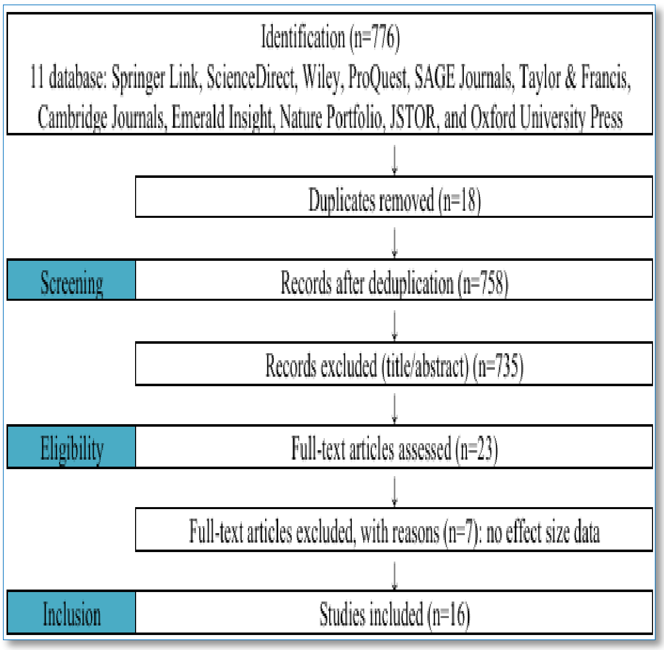


Figure 1. PRISMA flow diagram of study selection

Table 1. Citation, country, sample type, and intervention type

ID	Citation	Country	SampleType	nE	nC	InterventionType
A01	(Chen et al., 2023)	Taiwan	10; 11; 12	20	16	Gamification
A02	(Zulkefly & Yusof, 2025)	Malaysia	8	30	30	Gamification
A03	(Stasolla et al., 2025)	Italy	7; 8; 9	36	36	Gamification
A04	(Vekiri et al., 2024)	Cyprus	7; 8; 9; 10	69	27	Gamification
A05	(Silva Barbosa et al., 2024)	Brazil	10; 11; 12	94	39	Gamification
A06	(Egara & Mosimege, 2023)	Nigeria	10	45	41	Flipped Classroom
A07	(Fitrah et al., 2025)	Indonesia	12	46	45	Flipped Classroom
A08	(Tekin & Emmioğlu-Sarıkaya, 2020)	Turkey	10	34	33	Flipped Classroom
A09	(Song & Kapur, 2017)	Hong Kong	7	25	25	Flipped Classroom
A10	(Bhagat et al., 2015)	Taiwan	10; 11; 12	41	41	Flipped Classroom
A11	(Bolatli & Korucu, 2020)	Turkey	7	42	46	Flipped Classroom
A12	(Ramadhani et al., 2019)	Indonesia	11	33	29	Flipped Classroom
A13	(Zaitoun et al., 2023)	United Arab Emirates	11	25	25	Flipped Classroom
A14	(Ramadhani et al., 2020)	Indonesia	10	30	30	Flipped Classroom
A15	(Kirvan et al., 2015)	United States	7; 8	25	29	Flipped Classroom
A16	(Oyarinde, 2024)	Nigeria	11	93	82	Flipped Classroom

Table 2. Intervention duration, gamification, and pre-class materials

ID	Duration	Gamification Elements	Pre-class Materials
A01	1x	Points & scores; Leaderboards; Story/narrative; Feedback	Not applicable
A02	NA	Board game	Not applicable
A03	12w	Digital games; Missions & challenges; Feedback	Not applicable
A04	8w	Digital games; Missions and challenges; Story/narrative	Not applicable
A05	1x	Points & scores; Badges and trophies; Leaderboards; Player avatars; Visual themes	Not applicable
A06	4w	Not applicable	Instructional videos; Digital text materials; Digital flashcards; Practice exercises
A07	7w	Not applicable	Instructional videos; Digital text materials
A08	8w	Not applicable	Instructional videos
A09	2w	Not applicable	Instructional videos
A10	6w	Not applicable	Instructional videos
A11	3w	Not applicable	Instructional videos; Digital text materials; Digital flashcards; Practice exercises
A12	12w	Not applicable	Instructional videos; Digital text materials
A13	2w	Not applicable	Instructional videos; Worksheets
A14	8w	Not applicable	Digital text materials; Practice exercises
A15	2w	Not applicable	Instructional videos; Digital text materials; Worksheets
A16	NA	Not applicable	Computer Simulation

Note: 1x = one session; w = week(s); m = month(s); sem = semester; NA = not reported. "Not applicable" indicates that the characteristic does not apply to that intervention model.

The study quality assessment using the JBI checklist shows that most studies met the basic requirements of a quasi-experimental design, especially in the clarity of the cause–and–effect relationship (Q1) and the presence of a comparison group (Q2). The details of each study's achievement on items Q1–Q9 are presented in Table 3. However, several aspects still frequently appeared as weaknesses, particularly the equivalence of baseline conditions between groups (Q3), reporting the reliability of outcome measurements (Q7), and the completeness of data and continued participation of participants (Q8). Overall, based on the JBI assessment of the 16 studies, methodological quality fell into the moderate to high categories with scores of 5/9–9/9; as many as 7 studies were classified as high quality (7–9/9) and 9 studies were classified as moderate quality (5–6/9). This variation in design and reporting quality needs to be considered when interpreting the pooled effect sizes, especially when between-study heterogeneity is high.

Table 3. Methodological quality

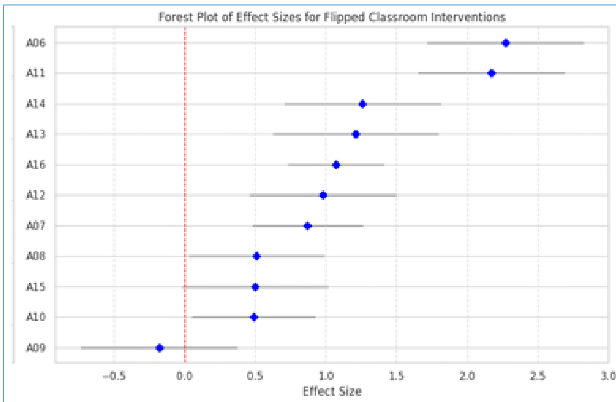
StudyID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Score (Y/9)	Quality
A01	Y	Y	N	Y	Y	Y	U	U	Y	6/9	Moderate
A02	Y	Y	U	U	Y	Y	U	U	Y	5/9	Moderate
A03	Y	Y	U	Y	Y	Y	U	U	U	5/9	Moderate
A04	Y	Y	U	U	Y	Y	U	U	Y	5/9	Moderate
A05	Y	Y	U	U	Y	Y	Y	U	Y	6/9	Moderate
A06	Y	Y	U	Y	Y	Y	Y	U	Y	7/9	High

StudyID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Score (Y/9)	Quality
A07	Y	Y	U	U	Y	Y	Y	U	Y	6/9	Moderate
A08	Y	Y	Y	U	Y	Y	Y	U	Y	7/9	High
A09	Y	Y	Y	U	Y	Y	U	U	Y	6/9	Moderate
A10	Y	Y	Y	Y	Y	Y	Y	U	Y	8/9	High
A11	Y	Y	Y	Y	Y	Y	U	U	Y	7/9	High
A12	Y	Y	N	Y	Y	Y	U	U	U	5/9	Moderate
A13	Y	Y	U	Y	Y	Y	Y	U	Y	7/9	High
A14	Y	Y	U	U	Y	Y	U	U	Y	5/9	Moderate
A15	Y	Y	Y	Y	Y	Y	Y	Y	Y	9/9	High
A16	Y	Y	Y	U	Y	Y	Y	U	Y	7/9	High

Effect sizes in the gamification studies ranged from -0.58 to 0.91, with a mean of 0.45. The distribution of effect sizes for each study and the pooled effect estimate are shown in Figure 2. The fixed-effect model produced a pooled effect size of g=0.335 (95% CI [0.122; 0.547]), which falls into the small to moderate category (Cohen, 1988). However, heterogeneity was very high (I^2=86.99%), indicating substantial differences across studies (Hegazi et al., 2025).

The random-effects model produced g=0.437 (95% CI [-0.162; 1.037]) and the confidence interval crossed zero. Thus, although the point estimate is positive, the overall effect of gamification cannot be considered statistically conclusive in this synthesis. One plausible contributor to this imprecision is study quality: several included studies show limitations in baseline comparability (Q3), measurement reliability reporting (Q7), and data completeness/attrition (Q8), which can inflate variability across effects and reduce confidence in the stability of the pooled estimate. In addition, variation in grade level, intervention duration, and the specific game elements used may also have contributed to the instability of the pooled estimate. This indicates that the impact of gamification is not yet consistent and is highly dependent on how it is implemented. Similar findings are reported in JPMS publications, where game-based learning in mathematics shows a positive impact on students' critical thinking and engagement, yet its effectiveness depends on how well game elements are aligned with instructional goals (Fatmawati & Siregar, 2026). This finding is in line with meta-analyses stating that the effect of gamification tends to be moderate but is highly sensitive to the design of game elements (Almeida et al., 2023). Long-term studies also show that the positive impact of gamification can diminish if competitive elements are too dominant (Diaz & Estoque-Loñez, 2024).

According to Self-Determination Theory, gamification will be more successful if its game elements make students feel they have choice and feel competent (Aubert et al., 2023). Conversely, if rewards or points are emphasized too strongly, students may become more focused on chasing scores than on truly understanding the material (Zhong et al., 2023). Therefore, differences in results across studies are to be expected because instructional designs are not the same.



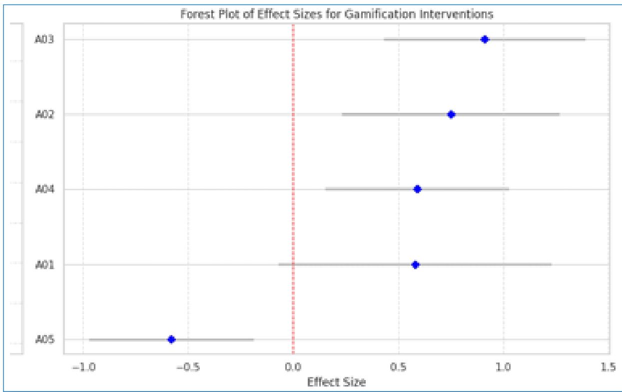
	StudyID	EffectSize	SE	CI_lower	CI_upper
0	A01	0.58	0.331662	-0.070058	1.230058
1	A02	0.75	0.264575	0.231433	1.268567
2	A03	0.91	0.244949	0.429900	1.390100
3	A04	0.59	0.223607	0.151731	1.028269
4	A05	-0.58	0.200000	-0.972000	-0.188000

Figure 2. Forest plot: effect sizes of the gamification intervention

Flipped classroom effect sizes ranged from -0.18 to 2.27, with a median of 0.98. The distribution of each study and the pooled effect estimate are shown in Figure 3. The fixed-effect model produced g=0.982 (95% CI [0.839; 1.126]), whereas the random-effects model produced g=1.009 (95% CI [0.619; 1.398]) despite high heterogeneity (I^2=86.07%).

These results are consistent with recent meta-analyses showing that the flipped classroom has a significant impact on academic outcomes across various fields, including mathematics (Deng et al., 2024; Güler et al., 2023). More advanced meta-analyses also show that its effectiveness tends to be stable when pre-class and in-class activities are designed to be well connected (Ma et al., 2025).

From the perspective of Cognitive Load Theory, separating learning before class and during face-to-face sessions helps regulate students' cognitive load more efficiently. Initial material can be designed according to students' capacity to process information, so that class time can be focused on practice and problem solving (Fischer et al., 2023). This pattern is appropriate within the scope of mathematics learning, because students often need time to understand concepts that are not immediately apparent. Supporting this perspective, JPMS studies indicate that well-designed instructional materials, such as interactive e-modules, can strengthen students' conceptual understanding, especially when combined with structured learning activities (Sitaresmi & Hanifah, 2026).



	StudyID	EffectSize	SE	CI_lower	CI_upper
0	A01	0.58	0.331662	-0.070058	1.230058
1	A02	0.75	0.264575	0.231433	1.268567
2	A03	0.91	0.244949	0.429900	1.390100
3	A04	0.59	0.223607	0.151731	1.028269
4	A05	-0.58	0.200000	-0.972000	-0.188000

Figure 3. Forest plot: effect sizes of the flipped classroom intervention

The two models differ in the strength and certainty of their pooled estimates. The flipped classroom showed a positive and statistically significant pooled estimate, whereas gamification showed a positive point estimate but with statistical uncertainty because the confidence interval crossed zero. At the same time, the very high heterogeneity in both models indicates that instructional effectiveness is highly dependent on context. The pooled effect estimates using the random-effects model with the HKSJ method along with heterogeneity indicators are presented in Figure 4. Recent educational meta-analyses also emphasize that differences in design and participant characteristics are often the main sources of heterogeneity (Pellegrini et al., 2025). Heterogeneity was also reflected in the between-study variance estimates (τ^2), with τ^2 = 0.451 in the overall model, τ^2 = 0.304 in the gamification subgroup, and τ^2 = 0.451 in the flipped classroom subgroup, indicating substantial dispersion in true effect sizes across studies. In the present review, plausible contributors to heterogeneity include differences in grade level, national and educational context, intervention duration, implementation intensity, and variation in the instructional components used within each model.

The use of a random-effects model helps estimate a more realistic average effect when results across studies vary greatly (Higgins et al., 2019). Even so, the average effect size cannot fully capture the complexity of implementation in real classroom conditions. Accordingly, the pooled estimates should be interpreted as average trends across diverse studies rather than as precise summary effects that can be expected to replicate uniformly across all secondary mathematics settings.

tau2	Q	p_Q	I2_%	note
0.451439	71.765584	2.020073e-11	86.065744	
0.303880	30.747566	3.446672e-06	86.990840	
0.450805	127.021466	0.000000e+00	88.190972	

Figure 4. Random-effects meta-analysis: HKSJ & heterogeneity

Subgroup analysis showed a significant difference between the flipped classroom and gamification ($p < 0.001$). The difference in mean effects between these models is shown in Figure 5 through the Q-subgroup (Q-between) test. However, meta-regression showed that model type was not a significant predictor after between-study differences were accounted for ($p = 0.146$). The meta-regression model also retained substantial residual heterogeneity (residual $\tau^2 = 0.451$), indicating that model type alone did not explain much of the between-study variance. As shown in Figure 6, subgroup analysis and meta-regression do not answer exactly the same statistical question. The subgroup test evaluates whether the pooled means of the two categories differ, whereas meta-regression evaluates whether model type explains variation in effect sizes after residual heterogeneity and estimation uncertainty are taken into account. When heterogeneity is high and the number of studies in each group is unbalanced, the power of the meta-regression test can decrease, so the moderator may appear non-significant even though the subgroup test shows differences. In this review, the subgroup sizes were also unequal ($k = 5$ for gamification versus $k = 11$ for the flipped classroom), which further reduces moderator-test stability and statistical power. Therefore, the subgroup result should be interpreted as an initial comparative signal rather than firm evidence that flipped classroom is inherently superior to gamification.

This finding is consistent with the instructional design literature emphasizing that implementation quality matters more than the model label (Kapur et al., 2022). Therefore, the effectiveness of mathematics learning innovations is highly dependent on design, especially the alignment between learning objectives, the required cognitive activities, and feedback mechanisms.

```
{'k_total': 16,
 'groups': ['Gamification', 'Flip Classroom'],
 'k_per_group': {'Flip Classroom': 11, 'Gamification': 5},
 'Q_total': 127.02146620137762,
 'Q_within_sum': 102.51314993316097,
 'Q_between': 24.508316268216646,
 'df': 1,
 'p': 7.39897939805445e-07,
```

Figure 5. Subgroup test (q-test): gamification vs flipped classroom

```
{'ref_group': 'Flip Classroom',
 'other_group': 'Gamification',
 'k': 16,
 'tau2': 0.45080503990838566,
 'difference_other_minus_ref': -0.5707350243636127,
 'SE': 0.3707670599907001,
 't': -1.5393358416951288,
 'df': 14,
 'p': 0.14601476927887422,
 'CI_low': -1.365951278949763,
 'CI_upp': 0.22448123022253763,
```

Figure 6. Meta-regression: gamification vs flipped classroom

The effectiveness of the flipped classroom tends to be higher when pre-class materials are presented in various formats such as text, images, and audio; this makes students actively process the material rather than only watching videos passively (Deng et al., 2024). In-class activities that emphasize collaborative problem solving and formative feedback have also been shown to have a tangible impact on learning outcomes (Güler et al., 2023).

In gamification, results tend to be more consistent when there is a balance between competition and activities that require thinking, compared with focusing only on rewards or points (Almeida et al., 2023). This indicates that mathematics learning will be more effective if it can combine conceptual understanding with student motivation. In line with this, JPMS studies highlight that integrating interactive media and meaningful learning activities can significantly enhance students' conceptual understanding and engagement in mathematics learning (Pangestuti et al., 2026). At the same time, these design differences help explain why heterogeneity remained very high across studies. The included interventions differed not only in their instructional labels, but also in their duration, intensity, learning materials, and classroom enactment. As a result, the meta-analytic findings should be interpreted as evidence of broad tendencies under diverse implementation conditions rather than as evidence of a single, context-free treatment effect.

Publication bias was assessed by visually examining the funnel plot and testing asymmetry using Egger's regression test and Begg's rank correlation test (Sadeghi, 2024). The distribution of studies in the funnel plot for the overall analysis is shown in Figure 7. In the overall analysis, the funnel plot appears broadly symmetrical, suggesting no obvious visual indication of publication bias; however, this visual impression should still be interpreted cautiously. The statistical test results were also not significant (Egger p=0.358; Begg p=0.314), and the summary is shown in Figure 8.

For the subgroup analysis, the results were also not significant, both for the flipped classroom (Egger p=0.6925; Begg p=0.4735) and for gamification (Egger p=0.1344; Begg p=0.8167). These results do not provide strong statistical evidence of funnel-plot asymmetry or small-study effects. Even so, these findings still need to be interpreted cautiously because tests such as Egger and Begg can be less sensitive when the number of studies is small, so an actual pattern may go undetected and appear as ordinary variation (Qiu et al., 2022). In particular, the gamification subgroup includes only k = 5 studies, which provides low statistical power for funnel-plot asymmetry tests. Therefore, non-significant Egger/Begg results should not be interpreted as evidence of no publication bias, but rather as inconclusive evidence under limited data.

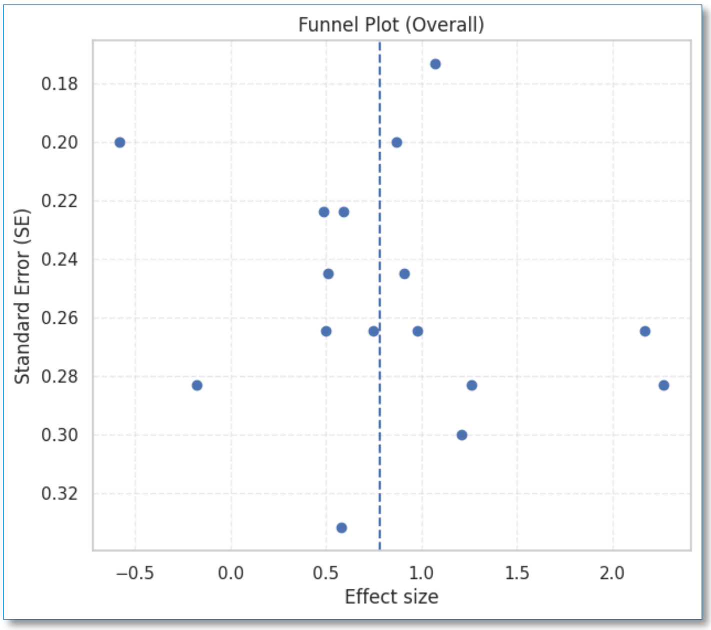


Figure 7. Publication bias

```
'Egger:'
{'k': 16,
 'egger_intercept': 4.088934062082272,
 'se_intercept': 4.305340787665372,
 't': 0.9497352854846945,
 'p': 0.35835491229644745,
 'note': 'p kecil => indikasi small-study effects/publication bias'}
'Begg :'
{'k': 16,
 'kendall_tau': 0.1932503014547343,
 'p': 0.31429216214942934,
```

Figure 8. Publication bias test egger & begg

Descriptively, the flipped classroom yielded a larger pooled estimate than gamification. However, this pattern should be interpreted cautiously. The pooled effects in this meta-analysis represent average trends across highly diverse studies rather than uniform effects that can be expected in all secondary mathematics settings. Given the very high heterogeneity, differences in implementation context, grade level, intervention duration, and instructional design are likely to play a substantial role in shaping outcomes. Moreover, although the subgroup analysis suggested a difference between the two approaches, the non-significant meta-regression result indicates that model type was not a robust explanatory factor once between-study variability was taken into account. Therefore, the present evidence does not justify a strong conclusion that flipped classroom is inherently superior to gamification. A more defensible interpretation is that the effectiveness of mathematics learning innovations depends primarily on instructional design quality and implementation context rather than on the instructional label alone.

The methodological quality of the selected studies was assessed using the JBI checklist, but several studies were still weak in terms of baseline equivalence between groups (Q3), reporting the reliability of instruments (Q7), and data completeness and the rate of participant attrition (Q8). In addition, the selection and appraisal processes were conducted by a single researcher, so the potential for rater subjectivity cannot be fully eliminated. Another limitation is the unequal number of studies across intervention groups, with only five gamification studies compared with eleven flipped classroom studies. This imbalance reduces the precision of subgroup comparisons and weakens the statistical power of moderator analyses, particularly under conditions of very high heterogeneity. Therefore, the results of this meta-analysis must be interpreted cautiously, particularly because between-study heterogeneity was very high, subgroup sizes were unequal, and study selection as well as quality appraisal relied on a single reviewer.

CONCLUSION

This study shows that both gamification and the flipped classroom can improve secondary students' mathematics achievement, but their effectiveness depends on how they are designed and used. Gamification shows a small-to-moderate positive effect, but the results are uncertain due to high variation across studies. In contrast, the flipped classroom shows a stronger and more consistent positive effect, although variation is also high. The difference between the two approaches appears significant at first, but further analysis suggests that the type of model alone does not determine effectiveness. Instead, learning outcomes are more influenced by the quality of instructional design and classroom implementation.

Therefore, teachers should focus on how the learning is designed. Flipped classroom approaches should connect pre-class and in-class activities and support active learning, while gamification should emphasize meaningful mathematical thinking rather than only rewards or competition. These findings should be interpreted with caution due to high heterogeneity, some methodological limitations, unequal study numbers, and single-reviewer screening. Future studies should use stronger research designs, clearer reporting, and more balanced comparisons.

REFERENCES

Almeida, C., Kalinowski, M., Uchôa, A., & Feijó, B. (2023). Negative effects of gamification in education software: Systematic mapping and practitioner perceptions. *Information and Software Technology*, *156*, 107142. https://doi.org/10.1016/j.infsof.2022.107142

Andrade, C. (2023). How to use percentiles to better understand standardized mean difference (SMD) as a measure of effect size. *Journal of Clinical Psychiatry*, *84*(4), 23F15028. https://doi.org/10.4088/JCP.23F15028

Asanre, A. A., Sondlo, A., & Adekunle, R. A. (2025). Investigating the interplay of social and biological factors on mathematics anxiety among senior secondary students. *Jurnal Pendidikan Matematika dan Sains*, *13*(2), 215–231. https://doi.org/10.21831/jpms.v13i2.84085

Aubert, A. H., Scheidegger, A., & Schmid, S. (2023). Gamified online surveys: Assessing experience with self-determination theory. *PLoS ONE*, *18*(10), e0292096. https://doi.org/10.1371/journal.pone.0292096

Bhagat, K. K., Chang, C. N., & Chang, C. Y. (2015). The impact of the flipped classroom on mathematics concept learning in high school. *Educational Technology & Society*, *19*(3), 134–142.

Bolatli, Z., & Korucu, A. T. (2020). Determining the academic achievement of students who use the flipped classroom method supported by a mobile application and their views on collaborative learning. *Bartın University Journal of Faculty of Education*, *9*(2), 229–251. https://doi.org/10.14686/buefad.631835

Caviola, S., Toffalini, E., Giofrè, D., Ruiz, J. M., Szűcs, D., & Mammarella, I. C. (2022). Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A meta-analysis on 906,311 participants. *Educational Psychology Review*, *34*(1), 1–38. https://doi.org/10.1007/s10648-021-09618-5

Cham, H., Lee, H., & Migunov, I. (2024). Quasi-experimental designs for causal inference: An overview. *Asia Pacific Education Review*, *25*(3), 1–15. https://doi.org/10.1007/s12564-024-09981-2

Chen, M. F., Chen, Y. C., Zuo, P. Y., & Hou, H. T. (2023). Design and evaluation of a remote synchronous gamified mathematics teaching activity that integrates multi-representational scaffolding and a mind tool for gamified learning. *Education and Information Technologies*, *28*(10), 13207–13233. https://doi.org/10.1007/s10639-023-11708-6

Demedts, F., Reynvoet, B., Sasanguie, D., & Depaepe, F. (2022). Unraveling the role of math anxiety in students' math performance. *Frontiers in Psychology*, *13*, 979113. https://doi.org/10.3389/fpsyg.2022.979113

Deng, R., Feng, S., & Shen, S. (2024). Improving the effectiveness of video-based flipped classrooms with question-embedding. *Education and Information Technologies*, *29*(10), 12677–12702. https://doi.org/10.1007/s10639-023-12303-5

Diaz, A. F., & Estoque-Loñez, H. (2024). A meta-analysis on the effectiveness of gamification on student learning achievement. *International Journal of Education in Mathematics, Science and Technology*, *12*(5), 1236–1253. https://doi.org/10.46328/ijemst.4185

Egara, F. O., & Mosimege, M. (2023). Effect of flipped classroom learning approach on mathematics achievement and interest among secondary school students. *Education and Information Technologies*, *29*(7), 8131–8150. https://doi.org/10.1007/s10639-023-12145-1

Fatmawati, R., & Siregar, T. J. (2026). Pengaruh model game-based learning terhadap kemampuan berpikir kritis matematis siswa pada materi bangun ruang sisi datar. *Jurnal Pendidikan Matematika dan Sains*, *14*(1), 131–140. https://doi.org/10.21831/jpms.v14i1.93601

Fischer, K., Sullivan, A. M., Cohen, A. P., King, R. W., Cockrill, B. A., & Besche, H. C. (2023). Using cognitive load theory to evaluate and improve preparatory materials and study time for the flipped classroom. *BMC Medical Education*, *23*(1), 1–12. https://doi.org/10.1186/s12909-023-04325-x

Fitrah, M., Sofroniou, A., Setiawan, C., Widihastuti, W., Yarmanetti, N., Jaya, M. P. S., Panuntun, J. G., Arfaton, A., Beteno, S., & Susianti, I. (2025). The impact of integrated project-based learning and flipped classroom on students' computational thinking skills: embedded mixed methods. *Education Sciences*, *15*(4), 1–20. https://doi.org/10.3390/educsci15040448

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

Hegazi, S., Aly, R., Yousif, D., Al-Nuaimi, S., Elshoeibi, A. M., Mahmoud, E., Elsayed, A. M., & Chivese, T. (2025). Prevalence of stroke in young adults in the Middle East and North Africa region: A systematic review and meta-analysis. *PLOS Global Public Health*, *5*(10), e0004666. https://doi.org/10.1371/journal.pgph.0004666

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Wiley-Blackwell.

Kapur, M., Hattie, J., Grossman, I., & Sinha, T. (2022). Fail, flip, fix, and feed – Rethinking flipped learning: A review of meta-analyses and a subsequent meta-analysis. *Frontiers in Education*, *7*, 1–19. https://doi.org/10.3389/feduc.2022.956416

Kirvan, R., Rakes, C. R., & Zamora, R. (2015). Flipping an algebra classroom: analyzing, modeling, and solving systems of linear equations. *Computers in the Schools*, *32*(3–4), 201–223. https://doi.org/10.1080/07380569.2015.1093902

Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: a meta-analysis and systematic review. *Educational Technology Research and Development*, *72*(2), 765–796. https://doi.org/10.1007/s11423-023-10337-7

Li, M., Ma, S., & Shi, Y. (2023). Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: a meta-analysis. *Frontiers in Psychology*, *14*, 1–17. https://doi.org/10.3389/fpsyg.2023.1253549

Ma, Y., Fan, H., & Ning, B. (2025). The effectiveness of the flipped classroom: A second-order meta-analysis. *International Journal of Educational Research*, *134*, 102855. https://doi.org/10.1016/j.ijer.2025.102855

Oyarinde, O. N. (2024). The impact of integrating computer simulation package in flipped classroom settings on students' mathematics achievement in Ondo State. *The Ethiopian Journal of Education*, *44*(2), 169–185. https://doi.org/10.2307/48815866

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 327, n71. https://doi.org/10.1136/bmj.n71

Pangestuti, A., Amarulloh, R. R., & Farizi, T. Al. (2026). Web-based Interactive Visual Media in Teaching Momentum and Impulse to Improve Students' Conceptual Understanding. *Jurnal Pendidikan Matematika dan Sains*, *14*(1), 11–24. https://doi.org/10.21831/jpms.v14i1.90237

Pellegrini, M., Day, E., Scarbrough, H. F., & Pigott, T. D. (2025). A meta-review of education meta-analyses: Relevance, applicability, and accessibility of findings. *AERA Open*, *11*. https://doi.org/10.1177/23328584251389562

Qiu, W., Fu, Y., Dang, Y., Cai, B., Tuo, S., Mao, B., Lin, R., Liu, Q., & Li, Y. (2022). Effectiveness and safety of weekly therapy versus 3-weekly therapy of paclitaxel plus carboplatin in women with ovarian cancer: A protocol of systematic review and meta-analysis. *BMJ Open*, *12*(3), e053182. https://doi.org/10.1136/bmjopen-2021-053182

Ramadhani, R., Bina, N. S., Sihotang, S. F., Narpila, S. D., & Mazaly, M. R. (2020). Students' critical mathematical thinking abilities through flip-problem based learning model based on LMS-google classroom. *Journal of Physics: Conference Series*, *1657*, 012025. https://doi.org/10.1088/1742-6596/1657/1/012025

Ramadhani, R., Umam, R., Abdurrahman, A., & Syazali, M. (2019). The effect of flipped-problem based learning model integrated with LMS-google classroom for senior high school students. *Journal for the Education of Gifted Young Scientists*, *7*(2), 137–158. https://doi.org/10.17478/jegys.548350

Sadeghi, M. (2024). Publication bias and funnel plot. In *Systematic Review and Meta-Analysis: Stepwise Approach for Medical and Biomedical Researchers*. Elsevier. https://doi.org/10.1016/B978-0-443-13428-9.00012-4

Schulz, A., Schürmann, C., Skipka, G., & Bender, R. (2022). Performing meta-analyses with very few studies. In E. Evangelou & A. A. Veroniki (Eds.), *Meta-research: Methods and protocols* (pp. 91–102). Springer. https://doi.org/10.1007/978-1-0716-1566-9_5

Silva Barbosa, J. F., Chalco Challco, G., & Bittencourt, I. I. (2024). Does gender-stereotyped gamification increase negative thinking? Results from an experimental study with logic tutoring systems. *Interactive Learning Environments*, *32*(9), 5632–5659. https://doi.org/10.1080/10494820.2023.2229145

Sitaresmi, P. D. W., & Hanifah. (2026). Desain e-modul interaktif untuk memfasilitasi pemahaman konsep peserta didik pada materi sistem persamaan linear dua variabel. *Jurnal Pendidikan Matematika dan Sains*, *14*(1), 107–120. https://doi.org/10.21831/jpms.v14i1.91702

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://doi.org/10.3929/ethz-b-000128354

Stasolla, F., Curcio, E., Borgese, A., Passaro, A., Di Gioia, M., Zullo, A., & Martini, E. (2025). Educational robotics and game-based interventions for overcoming dyscalculia: a pilot study. *Computers*, *14*(5), 201. https://doi.org/10.3390/computers14050201

Tekin, O., & Emmioğlu-Sarıkaya, E. (2020). Flipped classroom model in high school mathematics. *Bartın University Journal of Faculty of Education*, *9*(2), 301–314. https://doi.org/10.14686/buefad.559990

Tutal, Ö., & Yazar, T. (2021). Flipped classroom improves academic achievement, learning retention and attitude towards course: a meta-analysis. *Asia Pacific Education Review*, *22*(4), 655–673. https://doi.org/10.1007/s12564-021-09706-9

Vekiri, I., Meletiou-Mavrotheris, M., & Mannay, O. (2024). Using role models and game-based learning to attract adolescent girls to STEM. *Education Sciences*, *14*(8), 836. https://doi.org/10.3390/educsci14080836

Wang, Y., Wang, H., Wang, S., Wind, S. A., & Gill, C. (2024). A systematic review and meta-analysis of self-determination-theory-based interventions in the education context. *Learning and Motivation*, 87, 102015. https://doi.org/10.1016/j.lmot.2024.102015

Zaitoun, E., Shana, Z., Shater, A., Naser, K., & Mukattash, Z. (2023). Does flipping the classroom with videos and notetaking promote high school students' performance in mathematics? *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(6), em2274. https://doi.org/10.29333/ejmste/13200

Zajda, J. (2023). *The use of social constructivism to improve learning and performance*. In *Globalisation and inclusive schooling: Engaging motivational environments* (pp. 55–70). Springer. https://doi.org/10.1007/978-3-031-24400-1_5

Zhong, J., Wen, J., & Li, K. (2023). Do achievement goals differently orient students' academic engagement through learning strategy and academic self-efficacy and vary by grade. *Psychology Research and Behavior Management*, *16*. https://doi.org/10.2147/PRBM.S424593

Zulkefly, M. Z., & Yusof, F. M. (2025). Comparison of student achievement through circumference board gamification between schools in Putrajaya. *Malaysian Online Journal Of Educational Sciences*, *13*(3), 50–60. https://doi.org/10.22452/mojes.vol13no3.4

BIOGRAPHIES OF AUTHORS

Amelia Vega is a student in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang. Email: amelia.vega.2203118@students.um.ac.id

Dr. Erry Hidayanto, M.Si. is a lecturer in the Department of Mathematics, Universitas Negeri Malang, with a research focus on mathematics education and mathematical thinking. Email: erry.hidayanto.fmipa@um.ac.id

Dr. Makbul Muksar, S.Pd., M.Si. is a lecturer in the Department of Mathematics, FMIPA, Universitas Negeri Malang, with academic interests in applied mathematics. Email: makbul.muksar.fmipa@um.ac.id