Arshaka: Pedagogy and Learning Review
Volume 2, Number 1, 2026
E-ISSN : 3123-5794
Open Access: https://creativesmartvisionary.com/arshaka

The effectiveness of the flipped classroom on mathematical higher order thinking skills: a meta-analysis

Julham Hukom[1], Said Sudi Khassim[2]
[1] Department of Arabic Language Education, Universitas Negeri Makassar, Indonesia
[2] Department of Social Science and Humanities, The University of Dodoma, Tanzania

Correspondence: julham.hukom@unm.ac.id

Article Info

Article history:
Received Apr 27th 2026
Revised Jun 25th 2026
Accepted Jul 15th 2026

Keyword:
Flipped classroom; higher order thinking skills; mathematics; meta-analysis; moderator variable

DOI:
https://doi.org/10.65779/arshaka.v2i1.90

ABSTRACT

Higher order thinking skills (HOTS) are a core competency in 21st-century mathematics education, yet they remain difficult to achieve through conventional instruction. The flipped classroom has been proposed as an alternative that reverses the allocation of learning time so that higher order thinking activities can be facilitated more intensively in class. This study quantitatively synthesized the magnitude of the flipped classroom effect on mathematical HOTS through a meta-analysis. Eleven effect sizes drawn from experimental and quasi-experimental studies with a total of 1,091 participants were analyzed using a random-effects model with Hedges g as the effect size index. The pooled effect size was g = 0.82 (95% CI [0.59, 1.05]; $p < 0.001$), which is large, with moderate to substantial heterogeneity (I-squared = 67.6%). Publication bias tests indicated robust findings. Moderator analysis showed that education level, country, and flipped classroom variant acted as significant moderators, whereas the HOTS domain did not. These findings strengthen the evidence that the flipped classroom is effective for developing mathematical HOTS across contexts.

INTRODUCTION

The demands of contemporary mathematics education no longer stop at the mastery of procedures and facts, but have shifted toward the ability to reason, analyze, evaluate, and create. These abilities are summarized in the construct of higher order thinking skills (HOTS), which in the revised Bloom taxonomy occupy the highest cognitive levels, namely analyzing, evaluating, and creating (Anderson & Krathwohl, 2001). HOTS encompass critical thinking, problem solving, and creative thinking, and require students to apply knowledge to novel, non-routine situations rather than merely repeating memorized algorithms (Brookhart, 2010; Resnick, 1987).

Unfortunately, students' attainment of mathematical HOTS in many countries remains low. International assessments show that the mathematics performance of Indonesian students lies well below the average of OECD countries, including in higher order reasoning and creative thinking (OECD, 2023). One frequently cited cause is the dominance of teacher-centered instruction, which consumes most class time for delivering information and thus limits the space for practicing higher order reasoning (Akçayır & Akçayır, 2018). Under this pattern, the most cognitively demanding activities are precisely the ones students must complete independently at home, exactly when teacher support is unavailable.

The flipped classroom emerged as a response to this problem. Its central idea is to reverse the conventional sequence: the delivery of basic material is moved outside the classroom through videos, modules, or digital teaching materials, while face-to-face time is focused on collaborative activities, discussion, and the solving of complex problems that require the teacher's presence (Bergmann & Sams, 2012; Lage et al., 2000). Theoretically, moving lower-level cognitive load outside the classroom frees cognitive capacity for higher-level activities at the moment the most intensive support is available, an argument consistent with cognitive load theory and active learning (Abeysekera & Dawson, 2015; Bishop & Verleger, 2013).

Although many primary studies report a positive impact of the flipped classroom on learning outcomes, results across studies are not entirely consistent. Several meta-analyses have found positive effects ranging from small to moderate on general academic achievement (Bredow, Roehling, Knorp, & Sweet, 2021; Cheng et al., 2019; Strelanet al., 2020; van Altenet al., 2019), while a second-order synthesis reported a larger effect and highlighted high heterogeneity across fields and levels (Hew et al., 2021). Meta-analyses specifically targeting mathematics have likewise confirmed positive effects of the flipped classroom on general mathematical ability (Purnomo et al., 2022; Sulistyowati et al., 2023) and on mathematical HOTS (Pujiriyanto et al., 2024). In a related vein, blended learning strategies, which share the flipped classroom's emphasis on combining online and face-to-face components, have also demonstrated positive effects on students' mathematics learning achievement (Samritin et al., 2023; Setiawan et al., 2022). Affective factors such as self-efficacy and emotional intelligence have further been shown through meta-analytic evidence to be significantly associated with mathematics learning ability (Muhtadi et al., 2022; Muhtadi, Pujiriyanto, et al., 2022; Martaputri et al., 2021). Despite this growing body of meta-analytic evidence, differences in sample size, education level, the accompanying technology variant, and the HOTS domain measured make it difficult to draw a single conclusion by examining studies one at a time, so a quantitative synthesis that combines evidence across studies while tracing the sources of variation is needed (Borenstein et al., 2009).

Based on this background, the present study aimed to answer two main questions: (1) How large is the overall effect of the flipped classroom on mathematical HOTS? and (2) Which moderator variables explain the variability in effect sizes?

METHODS

Inclusion Criteria

A study was considered for analysis if it met all of the following criteria. First, the study was experimental or quasi-experimental, comparing a group taught with the flipped classroom (experimental group) and a group taught conventionally (control group). Second, the dependent variable measured was a higher order thinking skill in mathematics learning, covering one of critical thinking, problem solving, or creative thinking. Third, the study reported sufficient statistical data to compute an effect size, either as the means and standard deviations of both groups or as a significance value accompanied by sample sizes. Fourth, the study involved students from primary education up to higher education. Studies that did not provide quantitative data convertible into an effect size were excluded.

Literature Search and Screening

The literature search was conducted in scientific databases and the Google Scholar search engine by combining the keywords ("flipped classroom" OR "flipped learning") AND ("higher order thinking skills" OR "problem-solving" OR "critical thinking" OR "creative thinking") AND ("mathematics" OR "mathematical") AND ("experimental" OR "quasi-experimental"). Screening was carried out in stages following the PRISMA flow (Page et al., 2021). Studies that passed all screening stages proceeded to data extraction. At the end of the process, a set of primary studies containing 11 independent effect sizes ready for analysis was obtained.

Data Extraction

Data were extracted into an Excel worksheet recording study identity (study code, author, and year), methodological characteristics (experimental and control group sample sizes), the statistical data for computing the effect size (means and standard deviations, or significance values), and the moderator variables (country, education level, flipped classroom variant, and HOTS domain). For studies reporting means and standard deviations, the effect size was computed directly from these statistics. For studies reporting only a significance value and sample sizes, the effect size was obtained by converting the equivalent test statistic (Borenstein et al., 2009; Lipsey & Wilson, 2001). A single study report could yield more than one effect size when it compared different treatments or HOTS domains; each comparison was treated as a separate effect size.

Data Analysis

Data analysis was performed with the help of Comprehensive Meta-Analysis (CMA) version 3. All effect sizes were corrected to Hedges g to address estimation bias in small samples (Hedges & Olkin, 1985). Each study's effect size was expressed as Hedges g together with its variance and 95% confidence interval. Because the primary studies came from diverse contexts, levels, and treatments, the synthesis used a random-effects model that assumes the population effect may vary across studies, with the between-study variance component estimated following the DerSimonian and Laird (1986) procedure. Each study's weight was determined by the inverse of its variance, accounting for the between-study variation component (Borenstein et al., 2009). Heterogeneity was evaluated with the Q statistic, the I-squared index expressing the proportion of variation arising from genuine differences between studies, and the between-study variance estimate (tau-squared) (Higgins & Thompson, 2002). The interpretation of effect size magnitude followed Cohen's conventions, namely small (around 0.20), medium (around 0.50), and large (around 0.80) (Cohen, 1988).

Publication bias was examined through several complementary approaches: visual inspection of the funnel plot, the Egger regression test to detect asymmetry (Egger, Davey Smith, Schneider, & Minder, 1997), the fail-safe N to gauge the robustness of the findings against unpublished null studies (Rosenthal, 1979), and the trim-and-fill procedure to estimate potentially missing studies (Duval & Tweedie, 2000). To trace the sources of heterogeneity, a moderator-variable analysis was conducted through subgroup analysis on four variables, namely education level, HOTS domain, country, and flipped classroom variant; differences in effect sizes between subgroups were tested using the between-group Q statistic (Borenstein et al., 2009).

RESULTS

Characteristics of the Studies Meeting the Inclusion Criteria

Eleven effect sizes from studies meeting the inclusion criteria were analyzed, with a total of 1,091 participants (546 in the experimental groups and 545 in the control groups). The studies came from a variety of countries, namely Indonesia, Ethiopia, Malaysia, Jordan, Vietnam, and China, and spanned the primary school, junior high school, senior high school, and higher education levels. The flipped classroom variants studied varied, ranging from the pure flipped classroom to combinations with particular approaches and technologies such as inquiry, problem-based learning, GeoGebra, realistic mathematics education, three-dimensional virtual worlds, and deep learning. The HOTS domains measured covered critical thinking, problem solving, and creative thinking. Full characteristics together with each study's effect size are presented in Table 1. All effect sizes were positive, indicating that in every study the flipped classroom group outperformed the conventional group. Individual effect sizes ranged from 0.47 to 2.01, with the majority in the medium to large range.

Table 1 Characteristics of the Included Studies

Author (year)	Country	Level	FC Variant	HOTS Domain	nE	nC
Yohannes & Chen (2024)	Ethiopia	HE	RME + FC	Critical Thinking	32	32
Ariani et al. (2022a)	Indonesia	HE	Android module + inquiry + FC	Problem Solving	41	41
Ariani et al. (2022b)	Indonesia	HE	PPT module + FC	Problem Solving	41	41
Ariani et al. (2022c)	Indonesia	HE	Android module + inquiry + FC	Creative Thinking	41	41
Ariani et al. (2022d)	Indonesia	HE	PPT module + FC	Creative Thinking	41	41
Ramli et al. (2024)	Malaysia	HE	FC	Creative Thinking	25	24
Tabieh & Hamzeh (2022)	Jordan	SHS	3D Virtual Worlds + FC	Creative Thinking	180	180
Nguyen et al. (2023)	Vietnam	JHS	GeoGebra + FC	Problem Solving	37	37
Wei et al. (2020)	China	PS	FC	Problem Solving	44	44
Wijayanto et al. (2025)	Indonesia	PS	Deep Learning + FC	Problem Solving	16	16
Antara et al. (2026)	Indonesia	JHS	PBL + FC	Critical Thinking	48	48

Note. nE = experimental group sample size; nC = control group sample size; FC = flipped classroom; PS = primary school; JHS = junior high school; SHS = senior high school; HE = higher education.

Overall Effect Size

The random-effects synthesis produced a overall effect size of Hedges g = 0.82 with a 95% confidence interval of [0.59, 1.05]. This value differed significantly from zero ($z = 6.95$; $p < 0.001$) and is classified as a large effect under Cohen's (1988) conventions. In other words, students who learned through the flipped classroom on average attained markedly higher mathematical HOTS than students under conventional instruction. The heterogeneity test indicated meaningful variability between studies ($Q = 30.83$; $df = 10$; $p = 0.001$) with I-squared = 67.6%, signaling that about two-thirds of the variation in effect sizes stems from genuine between-study differences rather than mere sampling error (Higgins & Thompson, 2002). The 95% prediction interval of [0.06, 1.58] indicates that in almost all comparable contexts the effect remains positive. A summary of the parameters is presented in Table 2, while the distribution of effect sizes across studies is visualized in Figure 1.

Table 2 Summary of the Overall Effect Size and Heterogeneity

Parameter	Value
Number of effect sizes (k)	11
Total participants (experimental / control)	546 / 545
Overall effect size, Hedges g (random-effects)	0.82
Standard error (SE)	0.12
95% confidence interval	[0.59, 1.05]
Significance test (z; p)	6.95; < 0.001
95% prediction interval	[0.06, 1.58]
Heterogeneity test Q (df); p	30.83 (10); 0.001
I-squared index	67.6%
Tau-squared; Tau	0.098; 0.31
Fixed-effect estimate (comparator)	0.88

Note. The primary estimate uses the random-effects model; the fixed-effect estimate is provided as a comparator.

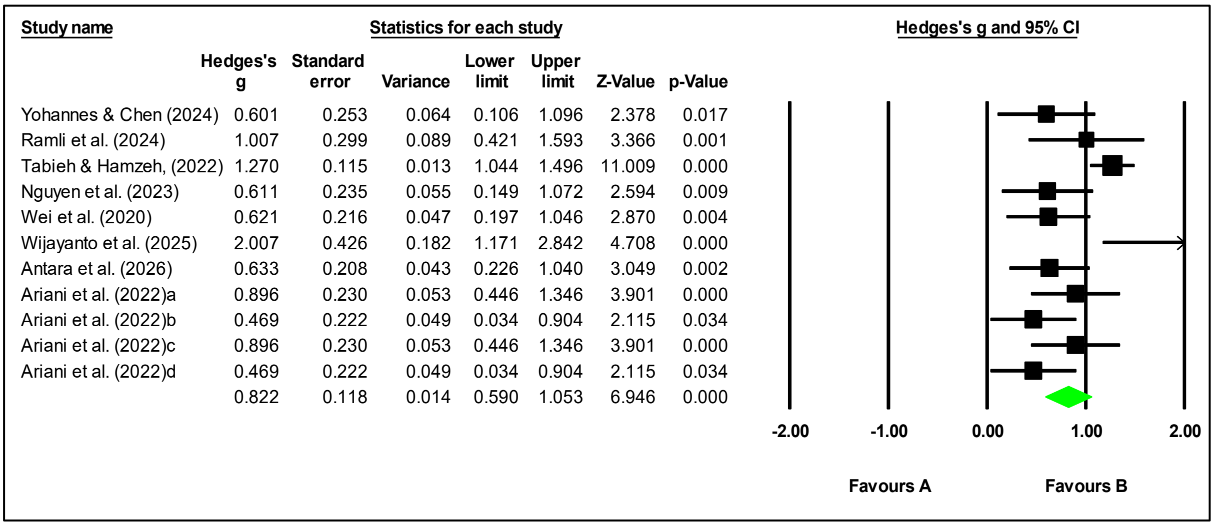

Figure 1 Forest Plot of Effect Sizes (Hedges g) Per Study and the Overall Effect Size

Publication Bias

Visual inspection of the funnel plot (Figure 2) showed a relatively symmetric distribution of studies around the pooled estimate, with no conspicuous gap on either side. This impression was reinforced by formal testing. The Egger regression test found no significant asymmetry ($t = -1.01$; $p = 0.34$) (Egger et al., 1997), as did the Begg rank correlation test ($tau = 0.13$; $p = 0.58$). Rosenthal's fail-safe N of 641 far exceeds the critical threshold of $5k + 10 = 65$, meaning that hundreds of unpublished null studies would be required to nullify the significance of the finding (Rosenthal, 1979). The trim-and-fill procedure imputed no missing studies (Duval & Tweedie, 2000). Overall, these indicators show that the conclusion of this meta-analysis is robust against the threat of publication bias. A summary is presented in Table 3.

Table 3 Results of the Publication Bias Tests

Method	Statistic	Result
Egger regression test (intercept)	t = -1.01; p = 0.34	Not significant
Begg rank correlation test	tau = 0.13; p = 0.58	Not significant
Rosenthal fail-safe N	641	>5k + 10 = 65
Trim-and-fill (imputed studies)	0	None

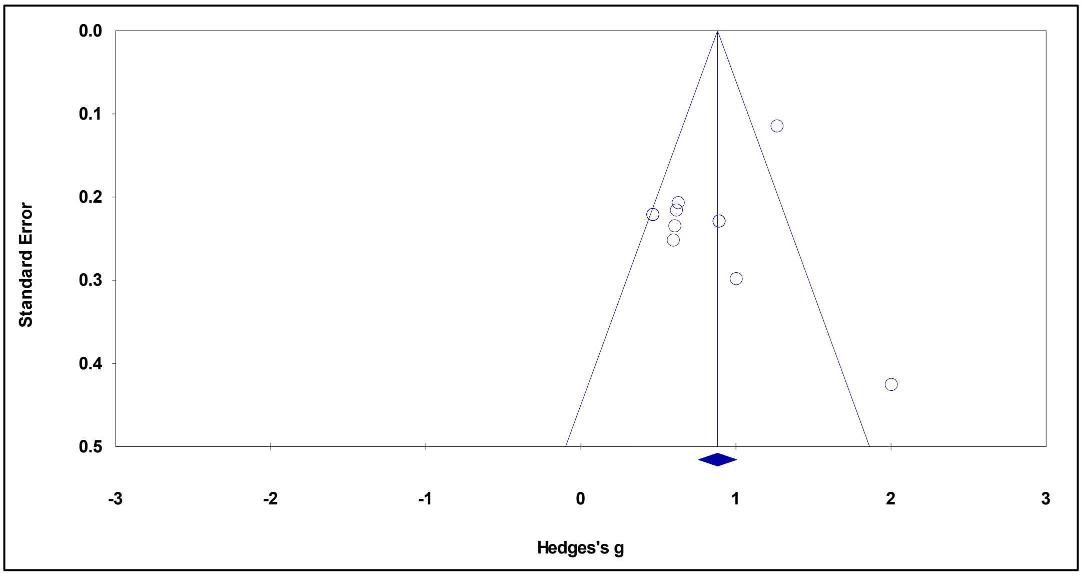

Figure 2 Funnel Plot Of Effect Size Against Standard Error

Moderator Variable Analysis

Subgroup analysis was conducted on four moderator variables to trace the sources of heterogeneity; full results are presented in Table 4. Education level proved to be a significant moderator (Q-between = 17.89; df = 3; p < 0.001): the largest effects were observed at the senior high school and primary school levels, whereas at the junior high school and higher education levels the effects were medium to large but more moderate. It should be noted that the estimate at the primary school level carried high internal heterogeneity (I-squared = 88.1%) and a wide confidence interval because it was supported by only two effect sizes.

Country also functioned as a significant moderator (Q-between = 14.96; df = 5; p = 0.011), as did the flipped classroom variant (Q-between = 29.65; df = 7; p < 0.001). However, most categories of these two moderators were represented by only a single effect size, so their comparisons are more descriptive than inferential. By contrast, the HOTS domain did not moderate the effect size (Q-between = 1.61; df = 2; p = 0.448): the effects on the critical thinking, problem solving, and creative thinking domains were relatively parallel at a large level. This finding suggests that the advantage of the flipped classroom is consistent across HOTS domains, while the variation in effects is more closely tied to the context of level and treatment design.

Table 4 Results of the Moderator Variable Analysis

Moderator Variable	k	Effect Size (g)	95% CI
Education level — Q-between = 17.89; df = 3; p < 0.001			
Higher education (HE)	6	0.70	[0.51, 0.89]
Senior high school (SHS)	1	1.27	[1.04, 1.50]

Moderator Variable	k	Effect Size (g)	95% CI
Junior high school (JHS)	2	0.62	[0.32, 0.93]
Primary school (PS)	2	1.27	[-0.09, 2.62]
HOTS domain — Q-between = 1.61; df = 2; p = 0.448			
Critical thinking	2	0.62	[0.31, 0.93]
Problem solving	5	0.81	[0.44, 1.19]
Creative thinking	4	0.93	[0.55, 1.32]
Country — Q-between = 14.96; df = 5; p = 0.011			
Indonesia	6	0.80	[0.48, 1.11]
Jordan	1	1.27	[1.04, 1.50]
Malaysia	1	1.01	[0.42, 1.59]
China	1	0.62	[0.20, 1.05]
Vietnam	1	0.61	[0.15, 1.07]
Ethiopia	1	0.60	[0.11, 1.10]
Flipped classroom variant — Q-between = 29.65; df = 7; p < 0.001			
Pure FC	2	0.76	[0.40, 1.12]
Android module + inquiry + FC	2	0.90	[0.58, 1.21]
PPT module + FC	2	0.47	[0.16, 0.78]
3D Virtual Worlds + FC	1	1.27	[1.04, 1.50]
GeoGebra + FC	1	0.61	[0.15, 1.07]
RME + FC	1	0.60	[0.11, 1.10]
PBL + FC	1	0.63	[0.23, 1.04]
Deep Learning + FC	1	2.01	[1.17, 2.84]

DISCUSSION

The main finding of this study confirms that the flipped classroom produces a large and significant effect on mathematical HOTS, with a pooled effect size of $g = 0.82$. This magnitude is large by Cohen's (1988) conventions and higher than the average effect reported by previous meta-analyses on general academic achievement, such as $g = 0.50$ in the cross-disciplinary synthesis by Strelan, Osborn, and Palmer (2020) and the range of $g = 0.20–0.53$ in the higher-education synthesis by Bredow et al. (2021). Notably, this finding is also consistent with prior mathematics-specific meta-analyses: Purnomo et al. (2022) reported $d = 0.73$ for the flipped classroom effect on general mathematical ability, Sulistyowati et al. (2023) found a similar positive pattern across multiple mathematics skill types, and Pujiriyanto et al. (2024) reported an overall effect size of $g = 1.06$ for the flipped classroom on mathematical HOTS. The present result thus falls within the range established by these prior syntheses while adding precision through its focus on studies that explicitly measured HOTS outcomes. This difference is reasonable given that the present study focuses on higher order thinking outcomes, which benefit most from reallocating class time to challenging cognitive activities; a similar pattern is seen in the second-order synthesis reporting a larger effect accompanied by high heterogeneity (Hew et al., 2021).

The result aligns with the theoretical foundation of this model. Moving the delivery of basic material outside the classroom frees face-to-face time for higher order reasoning activities that most require teacher guidance and peer interaction, a mechanism explained through cognitive load theory and active learning (Abeysekera & Dawson, 2015; Bishop & Verleger, 2013). When students confront non-routine problems at a time when support is optimally available, the processes of analyzing, evaluating, and creating can proceed more effectively (Anderson & Krathwohl, 2001; Brookhart, 2010). The consistency of effect direction across all studies, the prediction interval that remains positive, and the robustness of the findings against publication bias testing (Duval & Tweedie, 2000; Egger et al., 1997; Rosenthal, 1979) strengthen the conviction that this positive effect is not an artifact of selective publication.

The observed heterogeneity can be partly explained by moderator variables. Education level emerged as a significant moderator, with a tendency toward larger effects at certain school levels than at the higher education level. This finding complements Strelan et al.'s (2020) conclusion that the flipped classroom is beneficial across levels, while also showing that the magnitude of the benefit can differ between levels. This pattern can be understood from the perspective that students at earlier levels have greater room for improvement when conventional instruction is replaced with a more active model. Nevertheless, estimates for subgroups with a limited number of studies, particularly primary school, should be interpreted cautiously because they carry high internal heterogeneity and wide confidence intervals (Borenstein et al., 2009).

The flipped classroom variant and country also produced statistically significant between-subgroup differences. However, because most categories were represented by a single effect size, those differences are better read as a descriptive picture of the diversity of application contexts rather than strong evidence of the superiority of one variant over another. An interesting pattern is that flipped classrooms combined with particular technologies or active-learning approaches tended to yield strong effects, suggesting synergy between reversing the classroom structure and additional instructional support (Akçayır & Akçayır, 2018; Cheng et al., 2019). This observation resonates with the meta-analytic finding by Cahyani et al. (2024) that the integration of GeoGebra into the flipped classroom yielded a large effect on students' mathematics skills, and with evidence that blended learning strategies more broadly enhance mathematics achievement (Samritin et al., 2023; Setiawan et al., 2022). Furthermore, affective factors such as self-efficacy, which have been meta-analytically shown to relate positively to mathematics learning ability (Muhtadi, Assagaf, & Hukom, 2022), may partially explain why certain flipped classroom variants that promote student autonomy and confidence yield stronger effects. In contrast to these three moderators, the HOTS domain did not moderate the effect. The effectiveness of the flipped classroom was relatively equal in developing critical thinking, problem solving, and creative thinking, indicating that the core mechanism of the flipped classroom—reallocating class time to higher order cognitive activities, operates generally across types of thinking skills and is not limited to a particular domain (Anderson & Krathwohl, 2001; Brookhart, 2010).

Practically, these findings provide a strong empirical basis for educators and curriculum developers to adopt the flipped classroom as a strategy for developing mathematical HOTS. Because its positive effect is consistent across HOTS domains, the model can be applied without restricting it to a single type of thinking skill. Educators are advised to use face-to-face time for non-routine problem solving, discussion, and collaboration, while ensuring that the independent learning materials outside the classroom are well designed and accessible to students (Bergmann & Sams, 2012). The integration of culturally responsive approaches into mathematics instruction may further enhance the effectiveness of such models, as meta-analytic evidence suggests that culture-based mathematics learning instruction positively affects mathematical skills (Zuliana et al., 2025). Theoretically, this result enriches the evidence on the advantage of active learning based on reversing the classroom structure and underscores the importance of level context and treatment design in modulating the magnitude of the effect (Bredow et al., 2021; Purnomo et al., 2022; Strelan et al., 2020; Sulistyowati et al., 2023). For policymakers, integrating the flipped classroom needs to be accompanied by technological infrastructure support and the strengthening of teacher capacity, given that some effective variants combine the flipped classroom with particular technologies or approaches and given the implementation challenges that have been widely identified (Akçayır & Akçayır, 2018).

CONCLUSION

A meta-analysis of 11 effect sizes involving 1,091 participants showed that the flipped classroom produces a large and significant effect on mathematical higher order thinking skills (HOTS). This finding was robust against publication bias testing. Education level, country, and flipped classroom variant acted as moderators influencing the magnitude of the effect, whereas the HOTS domain did not, indicating that the advantage of the flipped classroom holds consistently across types of thinking skills. Accordingly, the flipped classroom deserves to be recommended as an effective instructional model for developing mathematical HOTS across various levels and contexts.

This study has several limitations. The number of effect sizes analyzed was relatively limited, and several moderator subgroups were represented by a single effect size, which constrains the strength of inference in certain subgroup analyses. In addition, some effect sizes were derived from significance values, so the estimates may be slightly less precise than those computed directly from means and standard deviations. Future research is encouraged to expand the coverage of primary studies, consider additional moderator variables such as intervention duration and implementation quality, and examine the long-term impact of the flipped classroom on mathematical HOTS.

REFERENCES

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research & Development, 34*(1), 1–14.

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education, 126*, 334–345.

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

Antara, I. K. J., Santyasa, I. W., Tegeh, I. M., Parwati, N. N., & Suartama, I. K. (2026). Effect of the problem-based learning flipped classroom model on junior high school students' digital literacy and critical thinking skills. *JKTP: Jurnal Kajian Teknologi Pendidikan*, *9*(2), 254-266.

Ariani, D. N., Sumantri, M. S., & Wibowo, F. C. (2022). The impact of android module-based inquiry flipped classroom learning on mathematics problem solving and creative thinking ability. *International Journal of Interactive Mobile Technologies*, *16*(24).

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

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. *Proceedings of the ASEE Annual Conference & Exposition.*

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009*). Introduction to meta-analysis.* John Wiley & Sons.

Bredow, C. A., Roehling, P. V., Knorp, A. J., & Sweet, A. M. (2021). To flip or not to flip? A meta-analysis of the efficacy of flipped learning in higher education. *Review of Educational Research, 91*(6), 878–918.

Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. ASCD.

Cahyani, I. A., Sujarwo, S., Imaroh, Y. R., Hukom, J., Yanuar, F. S., Martaputri, N. A., & Nisrina, N. (2024). Effectiveness of Geogebra integration into flipped classroom (GFC) on students mathematics skills: a meta-analysis study. *Jurnal Pendidikan Progresif*, *14*(3), 1754-1769.

Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research and Development, 67*(4), 793–824.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Lawrence Erlbaum Associates.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Hew, K. F., Bai, S., Huang, W., Dawson, P., Du, J., Huang, G., Jia, C., & Thankrit, K. (2021). On the use of flipped classroom across various disciplines: Insights from a second-order meta-analysis. *Australasian Journal of Educational Technology, 37*(2), 132–151.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558.

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education, 31*(1), 30–43.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Sage Publications.

Martaputri, N. A., Muhtadi, A. A., Hukom, J. A., & Samal, D. A. (2021). The corellation between emotional intelligence and academic achievement: a meta analysis study. *Jurnal Pendidikan Progresif*, *11*(3), 511-523.

Muhtadi, A., Assagaf, G., & Hukom, J. (2022). Self-efficacy and students' mathematics learning ability in Indonesia: A meta analysis study. *International Journal of Instruction*, *15*(3), 1131–1146. https://doi.org/10.29333/iji.2022.15360a

Muhtadi, A., Pujiriyanto, Syafruddin, K., Hukom, J., & Samal, D. (2022). A meta-analysis: Emotional intelligence and its effect on mathematics achievement. *International Journal of Instruction*, *15*(4), 745–762. https://doi.org/10.29333/iji.2022.15440a

Nguyen, A. T. T., Thanh, H. N., Le Minh, C., Tong, D. H., Uyen, B. P., & Khiem, N. D. (2023). Combining flipped classroom and GeoGebra software in teaching mathematics to develop math problem-solving abilities for secondary school students in Vietnam. *Mathematics Teaching Research Journal*, *15*(4), 69-97.

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71.

Pujiriyanto, Handaru, C., & Hukom, J. (2024). Meta-analysis of the effectiveness of the flipped classroom model on students' HOTS in mathematics. *Jurnal Teknologi Pendidikan*, *25*(3). https://doi.org/10.21009/jtp.v25i3.45848

Purnomo, B., Muhtadi, A., Ramadhani, R., Manaf, A., & Hukom, J. (2022). The effect of flipped classroom model on mathematical ability: a meta analysis study. *Jurnal Pendidikan Progresif*, *12*(3), 1201–1217. https://doi.org/10.23960/jpp.v12.i3.202216

Ramli, M. S., Ayub, A. M., Razali, F., & Ghazali, N. (2024). Effects of flipped classroom towards calculus performance and mathematical creative thinking skills among higher institution students. *Journal of Institutional Research South East Asia*, *22*(1), 70-96.

Resnick, L. B. (1987). *Education and learning to think*. National Academy Press.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641.

Samritin, S., Susanto, A., Manaf, A., & Hukom, J. (2023). A meta-analysis study of the effect of the blended learning model on students' mathematics learning achievement. *Jurnal Elemen*, *9*(1), 15–30. https://doi.org/10.29408/jel.v9i1.6141

Setiawan, A. A., Muhtadi, A., & Hukom, J. (2022). Blended learning and student mathematics ability in Indonesia: A meta-analysis study. *International Journal of Instruction*, *15*(2), 905–916. https://doi.org/10.29333/iji.2022.15249a

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.

Sulistyowati, E., Hukom, J., & Muhtadi, A. (2023). Meta-analysis of flipped classroom on students' mathematics abilities: effectiveness and heterogeneity analysis. *Jurnal Teknologi Pendidikan*, *25*(2), 140–159. https://doi.org/10.21009/jtp.v25i2.27433

Tabieh, A. A., & Hamzeh, M. (2022). The impact of blended-flipped learning on mathematical creative thinking skills. *Journal of Educators Online*, *19*(3), n3.

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review, 28*, 100281.

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.

Trisnawati, D., & Wijayanto, Z. (2025). The effect of flipped classroom model integrated with deep learning approach on students' mathematical problem-solving ability. *Jurnal Gentala Pendidikan Dasar*, *10*(2), 236-248.

Ulum, F., & Hukom, J. (2025). Flipped learning in foreign language learning in higher education: analysis of effectiveness and moderator variables. *Jurnal Pendidikan Progresif*, *15*(2), 1025-1040.

Yohannes, A., & Chen, H. L. (2024). The effect of flipped realistic mathematics education on students' achievement, mathematics self-efficacy and critical thinking tendency. *Education and Information Technologies*, *29*(13), 16177-16203.

Zuliana, E., Dwiningrum, S. I. A., Wijaya, A., & Hukom, J. (2025). The effect of culture-based mathematics learning instruction on mathematical skills: A meta-analytic study. *Journal of Education and Learning (EduLearn)*, *19*(1), 191–201.