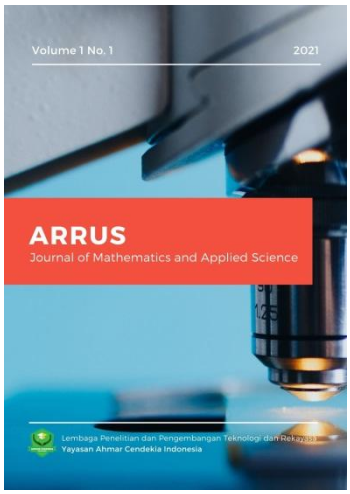
https://doi.org/10.35877/mathscience5554

OPEN ACCESS

ISSN : 2776-7922 (Print) / 2807-3037 (Online)



RESEARCH ARTICLE

Flipped Classroom, Home Learning Motivation, and Critical Thinking in Reaction Rate: A Quasi-Experiment

Asnawir

Madrasah Aliyah Swasta AlKhairaat Biromaru, Sigi, Sulawesi Tengah, Indonesia

*Corresponding author: Asnawir, Madrasah Aliyah Swasta AlKhairaat Biromaru, Sigi, Sulawesi Tengah, Indonesia.

E-mail: asnawirspd99@gmail.com

Abstract: This study examines whether a flipped classroom that combines a textbook with smartphone media affects home learning motivation and critical thinking in reaction rate among senior high school students. A nonequivalent control group design with pretest and posttest was applied to 51 Grade XI science students in two intact classes in Palu, Indonesia (experimental n = 25, control n = 26). Motivation was measured with a 20-item questionnaire based on the ARCS model, and critical thinking with a six-item essay test covering analysis, evaluation, and creation. Posttest scores were analysed with ANCOVA using the pretest as covariate, supported by Welch, Mann-Whitney, normalised gain, robust-error, bootstrap, and permutation analyses. Adjusted posttest motivation was 78.36 in the flipped class and 66.75 in the control class, $F(1, 48) = 122.52$, $p < .001$, partial $\eta^2 = .72$. Adjusted critical thinking was 46.78 and 36.46, $F(1, 48) = 326.55$, $p < .001$, partial $\eta^2 = .87$. Mean normalised gain was medium in the flipped class and low in the control class for both outcomes, and all six critical thinking indicators favoured the flipped class. The pooled correlation between the outcomes ($r = .44$) disappeared within classes. The flipped class scored higher on both outcomes, but the standardised differences are unusually large, one class served each condition, and reliability was not estimated, so the results are preliminary and should be treated as upper-range estimates until replicated with more classes.

Keywords: flipped classroom, reaction rate, critical thinking, learning motivation, quasi-experiment

1. Introduction

Reaction rate is one of the first quantitative topics in senior high school chemistry in which students must read tables and graphs, relate macroscopic observations to particle collisions, and decide which variable accounts for a change in speed. Recent work includes a meta-analysis of interventions that teach chemical kinetics (Çalik & Kurt, 2025), a case study of Grade 12 learners in which computer simulations with a predict-observe-explain sequence supported conceptual understanding (Jere & Mpeta, 2024), and studies of small-scale experiments with Grade 11 students (Supasorn et al., 2022) and of virtual versus physical experiments (Hunegnaw et al., 2025). A lecture-centred lesson, however, leaves little class time for the analysis, evaluation, and design tasks that kinetics problems require.

The flipped classroom moves first exposure to content out of the lesson and reserves class time for application. Syntheses report positive average effects, with gs of 0.20 to 0.53 across outcome categories in higher education (Bredow et al., 2021) and g = 0.53 across K–12 studies, although the K–12 prediction interval is wide (Li et al., 2025). In secondary

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

chemistry, a two-semester experiment in China found higher performance in a flipped class, with problem-solving gains appearing later (Yu et al., 2023), and an Indonesian study linked a flipped classroom with scientific literacy and critical thinking (Wijayanto et al., 2023). In many Indonesian schools, however, home study relies on a textbook and a smartphone rather than on a laptop or a full learning management system, and the effect of such a configuration on motivation and higher-order thinking in chemistry is not well documented, and secondary education in Central Sulawesi has itself been described in terms of technology adoption and pedagogical shifts (Ahmar & Azzajjad, 2025d). In the literature reviewed here, three gaps stand out. Few quasi-experimental studies in Indonesian senior high school chemistry measure home learning motivation and higher-order critical thinking in the same sample. Many studies analyse normalised gain alone, which can mislead when intact classes differ at pretest (Christman et al., 2024). The textbook-plus-smartphone configuration, common in schools with limited devices, has received little attention in reaction rate instruction.

This study therefore examined whether a flipped classroom with a textbook and smartphone media changed home learning motivation and critical thinking in reaction rate, relative to conventional teaching, and how the two outcomes were related. A nonequivalent control group design with 51 Grade XI science students was used, and posttest scores were compared after adjustment for pretest scores. The results favoured the flipped class on both outcomes, with differences that are much larger than those typically reported and that therefore call for cautious interpretation. The study addresses three research questions. RQ1: After adjusting for pretest scores, does home learning motivation differ between a flipped class and a conventionally taught class? RQ2: After the same adjustment, does critical thinking in reaction rate (analysis, evaluation, and creation) differ between the classes? RQ3: How are motivation and critical thinking related at posttest, between and within classes?

2. Literature Review

2.1. Flipped learning and learning outcomes

Meta-analytic evidence on flipped learning is generally positive but heterogeneous. Bredow et al. (2021) synthesised 317 higher education studies and found advantages for seven of eight outcome categories (gs = 0.20–0.53), and a methodological review discusses how design and reporting choices shape synthesised effects (Hew et al., 2021). Subject-specific syntheses report g = 0.263 for STEM higher education (Gong et al., 2024) and g = 0.402 for mathematics (Güler et al., 2023). A three-level meta-analysis of 129 K–12 studies reported g = 0.53 with a 95% prediction interval from −0.60 to 1.66 (Li et al., 2025). In chemistry, flipped designs have been tested with pre-service teachers (Aidoo et al., 2022) and in a Chinese secondary school, where performance gains lasted at least one year while problem-solving skills improved more gradually (Yu et al., 2023).

2.2. Motivation in flipped and mobile-supported learning

Flipped learning places more responsibility for preparation on the student, so home motivation matters. Self-determination theory links autonomy, competence, and relatedness to the quality of motivation (Ryan & Deci, 2020). A meta-analysis found a modest positive effect of flipped learning on motivation (Yazar & Tutal, 2023), a study of upper secondary students examined learning and motivation after flipped lessons (Ferriz-Valero et al., 2022), and autonomy support predicted motivation and engagement among 510 Cambodian Grade 12 students in a flipped English classroom (Mao et al., 2026). In Indonesian chemistry classrooms, motivation has been studied alongside software tools (Marpaung et al., 2021), video-based group tasks have been used in teacher education (Ahmar & Azzajjad, 2025a), and game-based and gamified STEAM tasks have been used to promote joyful learning and metacognitive thinking (Azzajjad & Ahmar, 2025; Ahmar & Azzajjad, 2025b). Where devices are limited, the medium for the home phase matters: Android-based e-learning applications have been built locally (Iskandar et al., 2023), motivation and

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

technology access shape test outcomes (Berlian et al., 2025), and engagement has been linked with outcomes in flexible formats (Parwati et al., 2025).

2.3. Critical thinking in science education

Critical thinking is a stated aim of science education, yet instructors differ in how they prioritise and practise it (Bellaera et al., 2021). Flipped designs have been linked with gains in critical thinking among 385 Grade 9 students, together with lower psychological stress (Atwa et al., 2022), among chemistry pre-service teachers (Aidoo et al., 2022), and in a study of scientific literacy (Wijayanto et al., 2023); another study explored the approach for enhancing critical thinking (Ma, 2023). Higher education work reported the greatest gains for a flipped format compared with online and in-class instruction (Orhan, 2023), a meta-analysis addressed IoT-based flipped classrooms (Syafruddin et al., 2023), and Indonesian studies connected flipped learning with self-efficacy (Dewi et al., 2024) and scientific attitude (Fuldiaratman et al., 2024). In chemical bonding, a correlational study examined critical thinking alongside a student behaviour variable (Ahmar & Azzajjad, 2025c), and problem-based and gamified designs have been examined for problem solving (Ahmar et al., 2023; Azzajjad et al., 2023). Few studies measure critical thinking with essay tasks that separate analysis, evaluation, and creation in reaction rate, and fewer still measure motivation in the same sample.

2.4. Hypotheses

Preparation outside class lets students meet new content at their own pace, which can raise confidence and motivation (Mao et al., 2026; Yazar & Tutal, 2023), and freed class time can be used for analysis, evaluation, and design, the processes that reaction rate problems involve (Aidoo et al., 2022; Atwa et al., 2022). The evidence, however, comes largely from higher education and other subjects, effect sizes vary widely (Hew et al., 2021; Li et al., 2025), and few studies measure both outcomes in one sample. The two outcomes are therefore treated as separate dependent variables and their association is examined explicitly.

The study tested the following hypotheses, summarised in Figure 1(a). The third is exploratory and follows from the expectation that more motivated students would also think more critically.

H1: Adjusted for pretest scores, home learning motivation is higher in the flipped class than in the conventional class.
H2: Adjusted for pretest scores, critical thinking in reaction rate is higher in the flipped class than in the conventional class.
H3: Posttest motivation and critical thinking are positively associated within classes (exploratory).

3. Research Method and Materials

3.1. Research design and participants

The study used a quasi-experimental nonequivalent control group design with pretest and posttest (Figure 1(b)). Two intact Grade XI science classes in a private senior high school in Palu City, Indonesia, took part in the 2024/2025 academic year. Classes were selected by purposive sampling and assigned to conditions without randomisation. The sample comprised 51 students, 25 in the experimental class (13 male, 12 female) and 26 in the control class (12 male, 14 female); student identity was replaced by codes before analysis. A sensitivity analysis indicated that, with these group sizes, $\alpha = .05$, and power of .80, the design could detect a standardised difference of at least 0.80 (Lakens, 2022); no a priori sample size justification was available because intact classes were used.

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

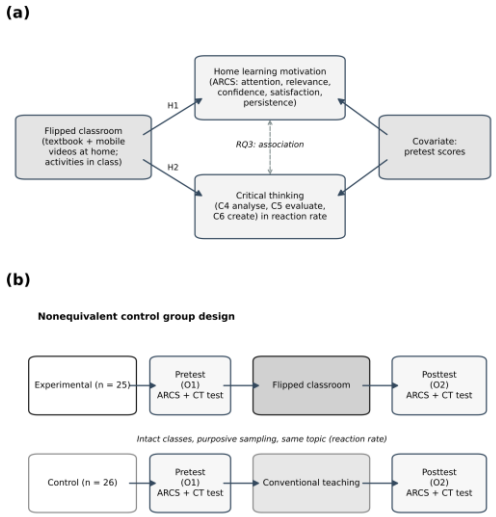


Figure 1: (a) Conceptual framework: the flipped classroom is expected to affect home learning motivation (H1) and critical thinking (H2) after adjustment for pretest scores, and the association between outcomes is examined as RQ3. (b) Nonequivalent control group design with pretest and posttest.

3.2. Instructional conditions

Both classes studied the same content, the factors that affect reaction rate (concentration, temperature, surface area, and catalyst) and their explanation through collision theory. In the flipped class, students studied the textbook and short videos or animations on their smartphones through Google Classroom or YouTube Edu before each lesson and completed a formative quiz; class time was used for group discussion, a simple practicum on rate factors, presentations, and teacher reinforcement of concepts, and reflection and enrichment tasks followed the lesson. In the control class, the teacher explained the content through lecture and questioning, followed by practice problems, and homework consisted of textbook exercises.

3.3. Instruments

Motivation was measured with a 20-item questionnaire adapted from Keller's ARCS model and modified to describe study at home with a textbook and smartphone. Items were rated from 1 (strongly disagree) to 5 (strongly agree), giving totals from 20 to 100, with four items for each of attention, relevance, confidence, satisfaction, and persistence. Critical thinking in reaction rate was measured with six essay items, two each at the analysis (C4), evaluation (C5), and creation (C6) levels of the revised Bloom taxonomy. Each item was scored from 0 to 10 with an analytic rubric, so total scores ranged from 0 to 60. Pretest and posttest used the same blueprint. Table 1 summarises both instruments. The recorded motivation totals contain decimals (for example, 63.7) and were analysed as recorded. Critical thinking totals differ from the indicator sums by at most 0.1 point, consistent with rounding.

Table 1: Instruments and content validity evidence

Instrument	Items and scoring	Structure	Score range	Mean expert rating	Mean Aiken's V
Home learning motivation questionnaire	20 Likert items (1–5)	Attention, relevance, confidence, satisfaction, persistence (4 items each)	20–100	4.55	0.89
Critical thinking test (reaction rate)	6 essay items, analytic rubric (0–10 each)	Analysis (C4), evaluation (C5), creation (C6) (2 items each)	0–60	4.60	0.90

Note. Ratings are means of two experts across 10 criteria per instrument (1–5 scale). Aiken's V was computed per criterion and averaged.

Two experts, one in chemistry education and one in learning assessment, rated each questionnaire and test criterion from 1 to 5. Mean ratings were 4.55 for the questionnaire

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

and 4.60 for the test, and mean Aiken's V values were 0.89 and 0.90 (Table 1). With only two raters, these values are descriptive and cannot be tested against conventional critical values. Internal consistency was not estimated because item-level responses are not part of the dataset analysed here, so content validity is the only instrument evidence available and the absence of reliability evidence is treated as a limitation. Current practice expects alpha to be reported for each scale (Taber, 2018; Lazenby et al., 2023), and recent Indonesian work has proposed two-tier diagnostic items in stoichiometry (Azzajjad et al., 2026).

3.4. Data analysis

Analyses were run in Python (SciPy and statsmodels). Baseline equivalence was checked with Welch t-tests and Mann-Whitney U tests, normality with Shapiro-Wilk tests, and variance equality with median-based Levene tests. The primary analysis was a one-way ANCOVA for each outcome, with posttest as dependent variable, class as factor, and pretest as covariate; homogeneity of slopes was tested through the class × pretest interaction. Because two outcomes were tested, α was set at .025 (Bonferroni), and a multivariate ANCOVA (Pillai's trace) served as an omnibus check. Welch t-tests were used for unadjusted comparisons because they control Type I error better when variances or group sizes differ (Delacre et al., 2017), with Mann-Whitney tests as distribution-free checks. Effect sizes were partial η^2, Hedges' g with 95% confidence intervals, and the rank-biserial correlation, interpreted with the education-specific cautions of Kraft (2023) and Funder and Ozer (2019). Because intact classes and one class per condition invite doubt about model-based inference, ANCOVA estimates were checked with HC3 standard errors, a percentile bootstrap (5,000 resamples), and a permutation test (2,000 permutations).

Normalised gain was computed per student as (posttest − pretest) / (maximum − pretest) (Hake, 1998) and classified as low (< .30), medium (.30–.70), or high (> .70). It served as a descriptive complement to the ANCOVA because it can reflect prior preparation (Christman et al., 2024). One control student scored lower on the motivation posttest than on the pretest, and that student's recorded N-gain was set to zero in the study file; this convention was kept,. Gain on each critical thinking indicator was analysed as raw posttest minus pretest with Holm correction across six tests, and within-class change was tested with paired t-tests and Wilcoxon tests. Associations used Pearson and Spearman correlations, pooled and within classes, and a partial correlation. Sex differences were examined exploratorily. The school granted permission, and students and parents or guardians gave consent.

4. Results and Discussion

4.1. Baseline equivalence and descriptive statistics

Motivation did not differ between classes at baseline (Welch $t(48.2) = 1.42$, $p = .162$, $g = 0.39$; Table 4). Critical thinking did: the experimental class started 2.44 points higher (Welch $t(48.6) = 2.24$, $p = .030$, $g = 0.62$). The classes were therefore not equivalent on critical thinking, which supports covariate adjustment rather than reliance on posttest differences or gain scores alone. The experimental class improved more (Table 2 and Figure 2). Every student improved on critical thinking, and 50 of 51 improved on motivation. Within-class changes were reliable in both classes (paired t-tests and Wilcoxon tests, all $p < .001$), so conventional teaching and repeated testing also raised scores, and only the between-class comparison speaks to the intervention.

Table 2: Descriptive statistics by class

Outcome	Class	N	Pretest M (SD)	Posttest M (SD)	N-gain M (SD)
Motivation (20–100)	Experimental	25	62.42 (10.11)	80.14 (10.25)	0.51 (0.18)
	Control	26	58.57 (9.23)	65.04 (8.69)	0.16 (0.08)
Critical thinking (0–60)	Experimental	25	28.54 (3.64)	47.81 (3.29)	0.62 (0.07)
	Control	26	26.10 (4.16)	35.46 (4.19)	0.28 (0.06)

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Note. M = mean; SD = standard deviation; N-gain = normalised gain (Hake, 1998). Motivation range 20–100; critical thinking range 0–60.

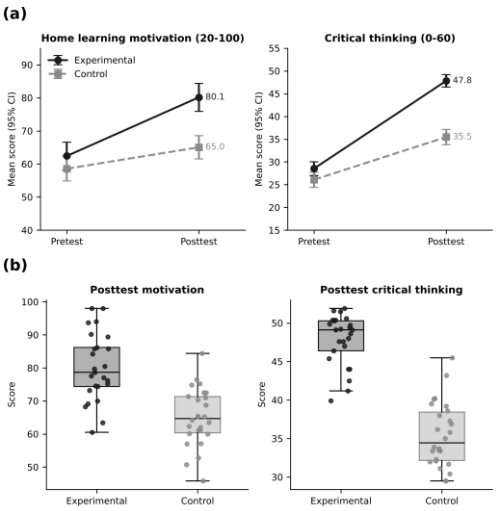


Figure 2: (a) Pretest and posttest means with 95% confidence intervals for motivation (left) and critical thinking (right). (b) Posttest score distributions; boxes show medians and interquartile ranges and dots are individual students.

4.2. Primary analysis: adjusted class differences

Assumption checks supported the ANCOVA. Residuals were consistent with normality (Shapiro-Wilk p = .979 for motivation and .638 for critical thinking), residual variances were similar across classes (Levene p = .162 and .758), and regression slopes did not differ between classes (class × pretest interaction p = .705 and .244). Some unadjusted distributions departed from normality or had unequal variances, which is why Welch and Mann-Whitney checks accompany the unadjusted comparisons in Table 4.

The multivariate test indicated an overall class effect after controlling for both pretests (Pillai's trace = .913, F(2, 46) = 240.70, p < .001). Univariate results are in Table 3. For motivation (RQ1, H1), the adjusted posttest mean was 78.36 in the experimental class and 66.75 in the control class, an adjusted difference of 11.61 points (95% CI [9.50, 13.72]). For critical thinking (RQ2, H2), the adjusted means were 46.78 and 36.46, a difference of 10.32 points (95% CI [9.17, 11.47]). Both hypotheses were supported at the Bonferroni-adjusted α of .025.

Table 3: ANCOVA of posttest scores with pretest as covariate

Statistic	Motivation	Critical thinking
Adjusted M, experimental	78.36	46.78
Adjusted M, control	66.75	36.46
Adjusted difference [95% CI]	11.61 [9.50, 13.72]	10.32 [9.17, 11.47]
F(1, 48)	122.52	326.55
p	< .001	< .001
Partial η^2	.72	.87
Covariate (pretest) slope	.91	.83
R^2	.91	.93

Note. Adjusted means are evaluated at the grand pretest mean (motivation 60.5; critical thinking 27.3).

The conclusions did not depend on model-based inference. With HC3 standard errors, the adjusted differences were 11.61 points (95% CI [9.41, 13.81]) for motivation and 10.32 points (95% CI [9.07, 11.56]) for critical thinking; percentile bootstrap intervals were [9.58, 13.73] and [9.21, 11.52], and permutation tests gave p < .001 for both outcomes. Removing the control student with a negative motivation gain left the adjusted motivation difference at 11.23 points (p < .001), and recomputing that student's N-gain as −0.07 left the Welch result unchanged. These checks address model assumptions, not the fact that one class served each condition.

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

4.3. Unadjusted comparisons, effect sizes, and normalised gain

Unadjusted comparisons agreed with the ANCOVA (Table 4 and Figure 3(a)). Posttest motivation differed by 15.1 points (Welch $t(47.1) = 5.67$, $g = 1.57$) and posttest critical thinking by 12.4 points ($t(47.2) = 11.73$, $g = 3.22$), and Mann-Whitney tests gave the same decisions (all $p < .001$). The pretest differences have small-to-moderate standardised size ($g = 0.39$ for motivation and 0.62 for critical thinking), which is why adjusted estimates are preferred over the raw g values for posttest and N-gain.

Table 4: Unadjusted comparisons between classes (experimental minus control)

Variable	Mean difference [95% CI]	Welch t (df)	p	Rank-biserial r	Hedges' g [95% CI]
Pretest motivation	3.85 [-1.60, 9.31]	1.42 (48.2)	.162	.23	0.39 [-0.16, 0.95]
Pretest critical thinking	2.44 [0.25, 4.64]	2.24 (48.6)	.030	.35	0.62 [0.05, 1.18]
Posttest motivation	15.10 [9.74, 20.46]	5.67 (47.1)	< .001	.74	1.57 [0.94, 2.20]
Posttest critical thinking	12.35 [10.23, 14.47]	11.73 (47.2)	< .001	.97	3.22 [2.38, 4.06]
Motivation N-gain	0.35 [0.27, 0.43]	8.68 (32.2)	< .001	.99	2.43 [1.70, 3.16]
Critical thinking N-gain	0.34 [0.30, 0.38]	18.15 (48.1)	< .001	1.00	5.02 [3.88, 6.15]

Note. Rank-biserial r = 1.00 indicates that every experimental N-gain exceeded every control N-gain.

Mean N-gain was 0.51 (medium) versus 0.16 (low) for motivation and 0.62 versus 0.28 for critical thinking; 25 of 26 control students had low motivation gain, and 14 of 26 had low critical thinking gain.

4.4. Indicator-level gains and sex

All six critical thinking indicators favoured the flipped class after Holm correction (all adjusted $p < .001$, $g = 1.76$–2.41). Raw gains in the experimental class ranged from 3.02 to 3.33 points per indicator and in the control class from 1.22 to 1.69. By domain, the mean gain difference was 3.59 points for analysis, 3.18 for evaluation, and 3.19 for creation, so creation did not show a larger advantage; because cognitive level is confounded with item content, differences among indicators should not be over-interpreted. Sex did not predict either outcome after adjustment (motivation $p = .472$, critical thinking $p = .257$), and the class × sex interaction was not significant ($p = .363$ and .663); with 12 to 14 students per cell this analysis is exploratory and has low power.

4.5. Association between motivation and critical thinking

Across all 51 students, posttest motivation and critical thinking correlated moderately ($r = .44$, $p = .001$; Spearman $\rho = .41$). Within classes the pattern disappeared (Figure 3(b)): $r = -.03$ ($p = .900$) in the experimental class and $r = -.48$ ($p = .014$) in the control class, a negative association in the latter. After controlling for both pretests and class, the partial correlation was $r = -.17$ ($p = .228$). The pooled correlation therefore reflects the class difference on both outcomes rather than a relation among students, and H3 was not supported. A weak link between classroom gamification and scientific literacy has likewise been reported in a correlational study of senior high school students (Ahmar et al., 2026). The flipped class may have raised both outcomes through shared activities rather than through motivation driving thinking, and narrow score spread leaves little room for within-class covariation.

4.6. Interpretation of the findings

After adjustment for pretest scores, students in the flipped class reported higher home learning motivation and scored higher on a critical thinking test about reaction rate than students taught conventionally. The direction agrees with a Chinese secondary chemistry experiment (Yu et al., 2023), Grade 9 results for critical thinking (Atwa et al., 2022), and a meta-analysis of motivation (Yazar & Tutal, 2023), although the present motivation difference is larger than that modest meta-analytic effect. Several features of the flipped condition could explain the differences, although this study cannot separate them. Home preparation with a textbook, short videos, and a quiz with immediate feedback fits the

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

ARCS aspects of confidence and satisfaction that the questionnaire measured. Class time was spent on discussion, a practicum, and presentations, activities that resemble the analysis, evaluation, and design tasks on the test, and such a close match between practice and assessment is a plausible contributor to the gains. Self-determination theory suggests that choice over when and how to prepare supports autonomy and competence (Ryan & Deci, 2020; Mao et al., 2026). These mechanisms remain interpretive because no process data were collected.

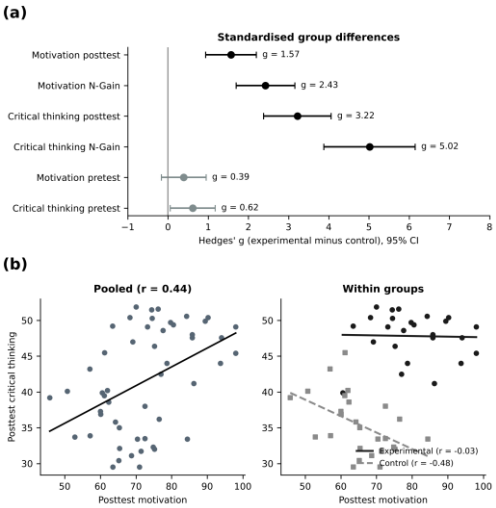


Figure 3: (a) Standardised group differences (Hedges' g with 95% confidence intervals); open markers are pretest differences and filled markers are posttest and N-gain differences. (b) Posttest motivation and critical thinking in the pooled sample (left) and within each class (right).

4.7. Size of the effects

The standardised differences in this study are far larger than usual. Syntheses of flipped learning report average effects in the small to moderate range, from g = 0.263 in STEM higher education (Gong et al., 2024) to g = 0.53 in K–12 (Li et al., 2025), whereas the posttest differences here were g = 1.57 for motivation and g = 3.22 for critical thinking, and the N-gain differences were g = 2.43 and g = 5.02. Kraft (2023) and Funder and Ozer (2019) urge caution in reading such effect sizes, and the present values call for it. No critical thinking N-gain overlapped between classes (Mann-Whitney U at its maximum, 650). Several design features could inflate them. Only one class served each condition, so class-level influences (teacher, timetable, peer climate) are confounded with the intervention, and the effective sample for the treatment contrast is two classes rather than 51 students. The test tasks were closely aligned with what the flipped class practised, so the test measured proximal rather than transferable skill. Scoring may not have been blind to class, and narrow within-class spread (SD = 3.29 for experimental critical thinking) magnifies standardised differences. These points do not show that the effect is spurious, only that its magnitude is uncertain and needs replication with more classes.

4.8. Practical implications and limitations

The conditions studied were modest: a textbook, a smartphone, a free platform, short videos, and a quiz. Teachers in schools with limited devices can adapt this configuration if every student can reach the media, because access can affect results independently of the construct (Berlian et al., 2025). Android applications (Iskandar et al., 2023) and simulations (Jere & Mpeta, 2024) could enrich the home phase, and monitoring completion of home tasks is advisable (Parwati et al., 2025).

Seven limitations qualify the findings. (1) Reliability was not estimated for either instrument, and content validity rests on two experts, so scores may contain measurement error that cannot be quantified here. (2) Assignment was not random, one class served each condition, and the classes differed on critical thinking at pretest; ANCOVA adjusts for the

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

observed difference but not for unmeasured class-level factors. (3) The sample is small (n = 51) and drawn from one private school in one city. (4) The pretest and posttest used the same blueprint, so testing effects may contribute to the large gains in both classes. (5) One control student's negative motivation gain was set to zero in the study file, a convention that slightly raises the control mean N-gain, although the sensitivity analysis showed no change in conclusions. (6) Intervention fidelity, teacher effects, and home-study completion were not measured. (7) Recorded totals contain decimals and small rounding differences, so item-level records should be archived. The effect sizes should be treated as upper-range estimates until replicated. A stronger test would randomise at least four classes per condition across two or more schools, use multilevel models, register the analysis plan, estimate reliability, use blind scoring, add a delayed posttest, and record process data.

5. Conclusion

This quasi-experiment examined whether a flipped classroom that combined a textbook with smartphone media changed home learning motivation and critical thinking in reaction rate among Grade XI science students. After adjustment for pretest scores, the flipped class scored higher than the conventionally taught class on both outcomes, with adjusted differences of 11.61 points for motivation and 10.32 points for critical thinking and partial η^2 of .72 and .87. The direction and the statistical decision held across unadjusted Welch tests, Mann-Whitney tests, normalised gain, heteroscedasticity-robust errors, bootstrap intervals, and permutation tests, and all six critical thinking indicators favoured the flipped class, with similar advantages for analysis, evaluation, and creation. Motivation and critical thinking were related in the pooled sample, but the relation disappeared within classes and after controlling for pretests and class, so the two outcomes should not be treated as products of one mechanism. Sex did not moderate either effect.

The findings must be read with restraint. The standardised differences exceed those reported in meta-analyses of flipped learning, only one class served each condition, the two classes differed on critical thinking at pretest, the test was closely aligned with the practised activities, and reliability could not be estimated from the available data. The study supports the plausibility of a low-cost flipped design for reaction rate rather than a firm estimate of its effect. For teachers, the practical message is to ensure that every student can access the home materials and to protect class time for data analysis, procedure evaluation, and experimental design. For researchers, the next step is a randomised, multi-class replication across schools, analysed with multilevel models, with reliability evidence, blind scoring, and a delayed posttest.

References

Ahmar, D. S., & Azzajjad, M. F. (2025a). Multiple representation strategy through team-based representational activities with creative exploration via learning through video on cell fundamentals. *Daengku: Journal of Humanities and Social Sciences Innovation*. https://doi.org/10.35877/454RI.daengku3906

Ahmar, D. S., & Azzajjad, M. F. (2025b). OPTIMIST-G (Optimization of metacognitive thinking skills through STEAM-based gamification) in chemical bonding learning. *Daengku: Journal of Humanities and Social Sciences Innovation, 5*(1), 121–128. https://doi.org/10.35877/454ri.daengku3748

Ahmar, D. S., & Azzajjad, M. F. (2025c). Studi korelasional antara playing victim dan keterampilan berpikir kritis pada materi ikatan kimia [Correlational study between playing victim and critical thinking skills on chemical bonding]. *Journal of Chemistry Sciences and Education, 2*(1), 29–38. https://doi.org/10.69606/jcse.v2i01.236

Ahmar, D. S., & Azzajjad, M. F. (2025d). Technology adoption and pedagogical shifts in Central Sulawesi's secondary education. *EDUTREND: Journal of Emerging Issues and Trends in Education, 2*(2), 72–85. https://doi.org/10.59110/edutrend.604

Ahmar, D. S., Azzajjad, M. F., Santoso, T., & Ismail, M. (2026). The weak link between classroom gamification and scientific literacy: A multi-method correlational study of senior high school

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

students in a green chemistry context. *Journal of Pedagogical Research.* https://doi.org/10.33902/jpr.202648842

Ahmar, D. S., Magfirah, M., Poba, D., & Azzajjad, M. F. (2023). Gamification learning integrated with local wisdom based on character education, is there an effect on problem solving ability? *Jurnal Penelitian Pendidikan IPA, 9*(9), 7614–7620. https://doi.org/10.29303/jppipa.v9i9.4968

Aidoo, B., Anthony-Krueger, C., Gyampoh, A. O., Tsyawo, J., & Quansah, F. (2022). A mixed-method approach to investigate the effect of flipped inquiry-based learning on chemistry students learning. *European Journal of Science and Mathematics Education, 10*(4), 507–518.

Atwa, Z., Sulayeh, Y., Abdelhadi, A., Jazar, H. A., & Eriqat, S. (2022). Flipped classroom effects on grade 9 students' critical thinking skills, psychological stress, and academic achievement. *International Journal of Instruction, 15*(2), 737–750. https://doi.org/10.29333/iji.2022.15240a

Azzajjad, M. F., & Ahmar, D. S. (2025). CHEM-KAHOOT interactive gamification to realise joyful and characterful chemistry learning. *SPIN Jurnal Kimia & Pendidikan Kimia, 7*(1), 62–70. https://doi.org/10.20414/spin.v7i1.13131

Azzajjad, M. F., Ahmar, D. S., & Ahmar, A. S. (2023). Optimizing the ability to handle toxic and hazardous waste through problem-based learning. *ARRUS Journal of Mathematics and Applied Science, 3*(2), 75–89. https://doi.org/10.35877/mathscience2209

Azzajjad, M. F., Ahmar, D. S., & Kilo, A. K. (2026). Two-tier diagnostics for evaluating submicroscopic representations in stoichiometry. *Jambura Journal of Educational Chemistry, 8*(1), 64–76. https://doi.org/10.37905/jjec.v8i1.36488

Bellaera, L., Weinstein-Jones, Y., Ilie, S., & Baker, S. T. (2021). Critical thinking in practice: The priorities and practices of instructors teaching in higher education. *Thinking Skills and Creativity, 41*, Article 100856. https://doi.org/10.1016/j.tsc.2021.100856

Berlian, Z., Isnaini, M., Farwati, R., Metafisika, K., Velina, Y., & Fitriyanni, V. (2025). Paper vs. digital assessments: Evaluating critical thinking on ecological issues in Indonesian madrasahs. *Journal of Applied Science, Engineering, Technology, and Education, 7*(1), 41–55. https://doi.org/10.35877/454RI.asci3785

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. https://doi.org/10.3102/00346543211019122

Çalik, M., & Kurt, S. (2025). The effect of interventions teaching chemical kinetics on students' academic performance: A meta-analysis study. *International Journal of Science and Mathematics Education, 23*(6), 1519–1544. https://doi.org/10.1007/s10763-024-10523-w

Christman, E., Miller, P., & Stewart, J. (2024). Beyond normalized gain: Improved comparison of physics educational outcomes. *Physical Review Physics Education Research, 20*(1), Article 010123. https://doi.org/10.1103/PhysRevPhysEducRes.20.010123

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test. *International Review of Social Psychology, 30*(1), 92–101. https://doi.org/10.5334/irsp.82

Dewi, F., Afrida, A., & Ekaputra, F. (2024). Implementation of whimsical application using project based learning model integrated flipped classroom in improving student self efficacy. *Journal of Applied Science, Engineering, Technology, and Education, 6*(1), 88–95. https://doi.org/10.35877/454RI.asci3026

Ferriz-Valero, A., Østerlie, O., Penichet-Tomas, A., & Baena-Morales, S. (2022). The effects of flipped learning on learning and motivation of upper secondary school physical education students. *Frontiers in Education, 7*, Article 832778. https://doi.org/10.3389/feduc.2022.832778

Fuldiaratman, F., Syahri, W., & Ekaputra, F. (2024). Application of learning module flipbooks with a STEAM approach to improve scientific attitudes and collaboration skills. *Journal of Applied Science, Engineering, Technology, and Education, 6*(1), 63–68. https://doi.org/10.35877/454RI.asci3024

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gong, J., Cai, S., & Cheng, M. (2024). Exploring the effectiveness of flipped classroom on STEM student achievement: A meta-analysis. *Technology, Knowledge and Learning, 29*(2), 1129–1150. https://doi.org/10.1007/s10758-023-09700-7

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

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

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Hew, K. F., Bai, S., Dawson, P., & Lo, C. K. (2021). Meta-analyses of flipped classroom studies: A review of methodology. *Educational Research Review, 33*, Article 100393. https://doi.org/10.1016/j.edurev.2021.100393

Hunegnaw, T., Hailegebreal, T. D., Getahun, D. A., & Atlabachew, M. (2025). Effect of virtual experiments compared to physical experiments on students' conceptual understanding of chemical kinetics concepts. *European Journal of STEM Education, 10*(1), Article 03. https://doi.org/10.20897/ejsteme/16261

Iskandar, A., Mansyur, M., Ahmar, A. S., Muliadi, M., & Rahman, A. (2023). Android-based e-learning application design in schools. *Journal of Applied Science, Engineering, Technology, and Education, 5*(1), 1–7. https://doi.org/10.35877/454RI.asci1643

Jere, S., & Mpeta, M. (2024). Enhancing learners' conceptual understanding of reaction kinetics using computer simulations: A case study approach. *Research in Science Education, 54*(6), 999–1023. https://doi.org/10.1007/s11165-024-10182-5

Kraft, M. A. (2023). The effect-size benchmark that matters most: Education interventions often fail. *Educational Researcher, 52*(3), 183–187. https://doi.org/10.3102/0013189X231155154

Lakens, D. (2022). Sample size justification. *Collabra: Psychology, 8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Lazenby, K., Tenney, K., Marcroft, T. A., & Komperda, R. (2023). Practices in instrument use and development in chemistry education research and practice 2010–2021. *Chemistry Education Research and Practice, 24*(3), 882–895. https://doi.org/10.1039/D2RP00275B

Li, S., Fu, W., Liu, X., & Hwang, G.-J. (2025). Effectiveness of flipped classrooms for K–12 students: Evidence from a three-level meta-analysis. *Review of Educational Research, 95*(5), 929–971. https://doi.org/10.3102/00346543241261732

Ma, Y. (2023). Exploration of flipped classroom approach to enhance critical thinking skills. *Heliyon, 9*(11), Article e20895. https://doi.org/10.1016/j.heliyon.2023.e20895

Mao, K., Ruos, D., & Em, S. (2026). The interplay of instructional styles, academic motivation, and engagement in an English language flipped classroom context. *Daengku: Journal of Humanities and Social Sciences Innovation*. https://doi.org/10.35877/454RI.daengku4961

Marpaung, D. N., Pongkendek, J. J., Azzajjad, M. F., & Sukirno, S. (2021). Analysis of student motivation using Chemsketch on hydrocarbon topic in SMA Negeri 2 Merauke. *Journal of Applied Science, Engineering, Technology, and Education, 3*(1), 69–73. https://doi.org/10.35877/454RI.asci105

Orhan, A. (2023). Comparing the effectiveness of online, flipped, and in-class critical thinking instruction on critical thinking skills and dispositions in higher education: Flipped classroom produces the greatest gains. *International Journal of Technology in Education, 6*(2), 238–259. https://doi.org/10.46328/ijte.376

Parwati, N. N., Suharta, I. G. P., Sudatha, I. G. W., Prameswara, P. B., & Halili, S. H. (2025). Optimization student learning outcomes through HyFlex learning based on student engagement. *Journal of Applied Science, Engineering, Technology, and Education, 7*(3), 427–440. https://doi.org/10.35877/454RI.asci4134

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Supasorn, S., Wuttisela, K., Moonsarn, A., Khajornklin, P., Jarujamrus, P., & Chairam, S. (2022). Grade-11 students' conceptual understanding of chemical reaction rate from learning by using the small-scale experiments. *Jurnal Pendidikan IPA Indonesia, 11*(3), 433–448. https://doi.org/10.15294/jpii.v11i3.36535

Syafruddin, S., Agustina, I., Jemmy, J., Komari, K., & Santosa, T. A. (2023). The effectiveness of IoT-based flipped classroom model on students' critical thinking skills: A meta-analysis. *Jurnal Penelitian Pendidikan IPA, 9*(10), 883–891. https://doi.org/10.29303/jppipa.v9i10.5265

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.

https://doi.org/10.35877/mathscience5554

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Wijayanto, N., Gani, A., Hasan, M., & Widowati, A. (2023). The effectiveness of flipped classroom on scientific literacy and critical thinking improvement. *Jurnal Pendidikan Kimia, 15*(2), 119–129. https://doi.org/10.24114/jpkim.v15i2.45523

Yazar, T., & Tutal, Ö. (2023). The effect of flipped learning on motivation: A meta-analysis. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 36*(1), 1–22. https://doi.org/10.19171/uefad.1169794

Yu, L., Li, Y., Lan, Y., & Zheng, H. (2023). Impacts of the flipped classroom on student performance and problem solving skills in secondary school chemistry courses. *Chemistry Education Research and Practice, 24*(3), 1025–1034. https://doi.org/10.1039/D2RP00339B

This open access article is distributed under a Creative Commons Attribution (CC-BY-NC) 4.0 license.