Video-mediated Peer Explanation in Flipped Chemistry: A Randomized Experimental Study with Philippine Grade 11 STEM Students

Noe T. Hermosa [1,*] and Nenia Rose S. Hermosa [2]

[1] College of Teacher Education, Central Philippines State University—San Carlos Campus, Negros Occidental, Philippines
[2] Graduate School, State University of Northern Negros, Negros Occidental, Philippines
Email: noe_sky05@yahoo.com (N.T.H.); neniarose_s@yahoo.com (N.R.S.H.)
*Corresponding author

Manuscript received August 11, 2025; revised August 28, 2025; accepted November 10, 2025; published July 15, 2026

***Abstract*—This study examined the effectiveness of a "Flipped Application" instructional model—combining pre-class microlecture videos with Flip application-mediated peer discussion and in-class active problem solving—on Grade 11 chemistry performance in the Philippines. 70 Science, Technology, Engineering, and Mathematics (STEM) students were randomly assigned to an experimental group (Flipped Application) or a control group (traditional lecture and practice) and completed identical 25-item multiple-choice pre- and post-tests aligned to five mandated fourth-quarter competencies (structural isomerism; basic organic reactions; polymer formation and properties; and biomolecules). Baseline equivalence between groups was established. Within-group analyses using Wilcoxon signed-rank tests showed statistically significant gains for both the experimental group ($r = 0.86$) and the control group ($r = 0.87$), and a Mann-Whitney U test of post-test performance indicated that the Flipped Application group outperformed the control group with a large between-group effect ($r = 0.68$); parametric tests using percentage scores (0–100 scale) produced substantively similar conclusions. Findings suggest that short-cycle flipped instruction, when coupled with structured peer explanation and iterative feedback, can accelerate mastery of targeted chemistry competencies. However, the intervention covered only a seven-day sequence comprising two assessment days and five core instructional sessions. It used the same test at pre- and post-assessment, which may have inflated observed effect sizes due to novelty and practice effects. The study contributes classroom-based experimental evidence from a Philippine senior high school context. It offers practical guidance for integrating low-cost, video-based peer discourse in active-learning chemistry instruction.**

***Keywords*—flipped classroom, Flip application, peer instruction, chemistry education, Philippine Science, Technology, Engineering, and Mathematics (STEM) education, experimental design**

I. Introduction

International large-scale assessments have repeatedly shown that Filipino learners struggle to meet basic proficiency targets in science. In the Programme for International Student Assessment (PISA) 2018 cycle, Filipino 15-year-olds obtained an average science literacy score of 357, which was 132 points below the OECD mean of 489, and only about one in five students reached baseline proficiency (Level 2) [1]. The PISA 2022, roughly one in five (≈16–23%, depending on reporting metric) of Philippine students attained Level 2 or higher in science, compared to about three in four in OECD systems [2]. The Trends in International Mathematics and Science Study (TIMSS) 2019 similarly reported mean science scores among Filipino fourth graders far below the international average, signaling persistent weaknesses in conceptual understanding and scientific reasoning [3]. These findings point not only to content gaps but also to gaps in students' ability to explain and communicate scientific ideas clearly [4].

Chemistry at the Grade 11 level is particularly challenging in the Philippine K-12 curriculum because it requires students to integrate abstract molecular concepts with symbolic representations and apply them to real-world contexts late in the academic year (fourth quarter). Traditional lecture-based delivery often prioritizes content coverage over active sense-making, leaving students with fragmented procedural knowledge and a limited ability to articulate their reasoning.

Flipped classroom models attempt to address this problem by shifting the first exposure to new content outside the classroom (through short instructional videos) and devoting class time to higher-order problem-solving, feedback, and clarification [5–7]. In chemistry, flipped approaches have been associated with higher exam performance, reduced failure rates, and greater engagement [8, 9]. Meta-analyses generally report small-to-moderate average gains in achievement and satisfaction in secondary and postsecondary Science, Technology, Engineering, and Mathematics (STEM) contexts, and "flipped" is not just about watching videos [6, 7, 10–12]. Effective flipped instruction typically includes embedded formative assessment, opportunities for peer explanation, and explicit instructor scaffolding during class [5, 11, 12]. Socially shared regulation—students articulating their understanding to peers and responding to targeted prompts—appears to strengthen both cognitive processing and oral communication skills, which are crucial in science [13, 14]. Video-based discussion platforms, such as the Flip application, have been used to promote reflective explanation, peer feedback, and science communication skills in STEM settings, suggesting a possible means to integrate conceptual learning and communication practice [13, 15].

Building on these strands, the present study evaluates a "Flipped Application" model in Grade 11 chemistry. The model combines (a) pre-class micro lecture videos and guided reflection prompts, (b) peer video responses on Flip Application, and (c) in-class collaborative problem solving facilitated by the teacher. This design is grounded in social constructivist and self-regulated learning perspectives: students actively construct meaning before class, explain that

doi: 10.18178/ijiet.2026.16.7.2643

meaning to peers, and then refine it through guided feedback and application.

This study addresses four research questions:

(1) To what extent does the Flipped Application intervention produce a statistically significant improvement in Grade 11 chemistry students' performance, as measured by pre–post gains within the experimental group?
(2) To what extent does traditional (non-Flipped Application) instruction produce a statistically significant improvement in Grade 11 chemistry students' performance, as measured by pre–post gains within the control group?
(3) Is there a statistically significant difference in post-test chemistry scores between students who experienced the Flipped Application intervention and those who received traditional instruction?
(4) How do the magnitudes of these effects compare to typical effects reported for flipped instruction in prior literature?

II. Related Work and Theoretical Framework

A. Flipped Classroom and Chemistry Achievement

A substantial body of research has found that flipped classrooms can enhance conceptual understanding and improve course grades in chemistry. For example, documented achievement gains in organic chemistry courses after re-structuring lectures as pre-class videos plus in-class application [9]. Study reported that low-barrier flipped modules in large-enrollment general chemistry courses increased passing rates and reduced failure rates [8]. Meta-analyses in science and health professions education similarly report statistically reliable, though typically modest, average effects favoring flipped formats over traditional lecture [6, 7, 10–12].

These studies suggest that flipping can be an effective approach. However, the mechanisms matter: gains are strongest when out-of-class content is concise and aligned with in-class problem-solving, and when class time is used for formative assessment, explanation, and instructor guidance rather than passive review [5, 16].

B. Engagement, Self-regulation, and Peer Explanation

Recent work emphasizes that flipped classrooms may increase student engagement and self-regulated learning by requiring students to prepare before class and come ready to solve problems collaboratively [10]. When learners are prompted to generate explanations, identify misconceptions, or ask questions in advance, they arrive in class with specific gaps that can be addressed immediately. This targeted activation of prior knowledge can reduce cognitive overload during live instruction and enable deeper in-class practice [17].

C. Flip Application and Oral Scientific Communication

Flip application (a short-form video discussion platform) allows students to record brief responses to instructor prompts, watch peers' responses, and offer feedback. Studies in STEM and communication-intensive courses suggest that the Flip application can (i) increase students' willingness to articulate reasoning verbally, (ii) provide low-stakes rehearsal of disciplinary language, and (iii) promote peer-to-peer clarification of misconceptions [13–15]. These findings are especially relevant for Philippine learners, where national assessments have highlighted weaknesses not only in content mastery but in scientific explanation and argumentation [1, 4].

D. Research Gap and Contribution

While flipped classrooms and video-based peer discussion have both been studied, there is limited experimental evidence in Philippine senior high school chemistry that (a) integrates structured pre-class micro lectures with (b) Flip application-mediated peer reflection and (c) in-class collaborative problem solving led by the same teacher for all groups. The present study does not claim that combining "Flip application + flipped classroom" is unprecedented. Rather, its contribution is contextual and empirical, providing randomized evidence from Grade 11 STEM students in the Philippines, focusing on mandated fourth-quarter competencies over a short, well-defined instructional period. It also documents implementation fidelity and provides a mechanism-oriented interpretation of why the approach may have been effective.

III. Methods

We employed a randomized pre-test–post-test control-group design to estimate the short-term causal effect of the Flipped Application model on chemistry achievement [18].

A. Participants and Context

The sampling frame included all Grade 11 STEM students enrolled at one public senior high school during the 2023–2024 academic year. Eligibility criteria required (i) at least 80% attendance in chemistry, (ii) no prior experience with flipped or Flip application-based instruction, and (iii) no advanced coursework beyond the standard Grade 11 chemistry curriculum. Students who had missed more than 3 of the 5 core instructional sessions during the intervention window or had already completed advanced organic chemistry modules were excluded. The seven-day intervention window consisted of two assessment days (pre-test and post-test) and five core instructional sessions on the target fourth-quarter competencies.

70 students met the inclusion criteria and consented/assented to participate. This exceeds the minimum total sample size of 53 estimated via an a priori G*Power 3.1 analysis for detecting a medium effect ($r = 0.50$) using a two-tailed Mann–Whitney U test at $\alpha = 0.05$ and power = 0.80.

B. Randomization Procedure

To promote baseline equivalence, all eligible students first completed a 25-item chemistry pre-test aligned to the fourth-quarter learning competencies. Pre-test total scores were rank-ordered and split into three performance strata (lower, middle, upper). Within each stratum, students were individually assigned, using computer-generated random numbers, to either the Experimental group (Flipped Application; $n = 35$) or the Control group (Traditional instruction; $n = 35$).

After the assignment, students in each condition met in separate sections to accommodate scheduling practicality.

Crucially, both sections were taught by the same chemistry teacher using condition-specific lesson plans. Because students were randomized at the individual level before being organized into sections, clustering by intact pre-existing classes was minimized.

C. Instructional Conditions

1) Experimental group: Flipped application condition

Pre-class micro lectures: Each afternoon, the teacher uploaded a 5–7 mins video covering the next day's target concept, such as structural isomerism, functional groups, and basic organic reactions (combustion, addition, condensation, saponification), polymer formation and structure–property relationships, and macromolecules (proteins, nucleic acids, lipids, carbohydrates).

Flip application prompts: Students recorded short Flip application responses to prompts such as:

- "Explain today's concept in your own words and give one authentic example".
- "Identify a misconception you had and explain how it changed,"
- "Ask one question you still have and suggest how you might answer it".

This required students to verbalize chemical reasoning and anticipate peer feedback.

Downloadable pre-class materials: Concept maps and guided worksheets were posted to prepare students for in-class work.

In-class activities:

- The teacher highlighted representative Flip application responses (addressing misconceptions immediately).
- Peer instruction: students paired up to compare explanations before whole-class synthesis.
- Targeted mini-lecture and guided practice using slides and worked examples directly tied to the pre-class content.
- Small-group application tasks (e.g., linking polymer properties to structure, analyzing everyday products in terms of functional groups) and short formative quizzes.

2) Control group: Traditional instruction condition

- No pre-class video preparation or Flip application interaction.
- During class, the teacher (same teacher as the experimental group) used lecture and PowerPoint to introduce new content on the same curriculum competencies listed above.
- Guided practice occurred in class, followed by brief end-of-period quizzes or problem sets.
- Homework was assigned for additional practice, but no structured peer video reflection or pre-class activation was used.

D. Achievement Test

The primary outcome was a 25-item, four-option multiple-choice test. Items were evenly distributed across five competencies:

(1) Structural isomerism (5 items)
(2) Fundamental organic reactions (combustion, addition, condensation, saponification) (5 items)
(3) Polymer formation and structure (5 items)
(4) Relating polymer properties to structure (5 items)
(5) Structures and functions of biomolecules (proteins, nucleic acids, lipids, carbohydrates) (5 items)

Each correct answer was scored as 1 point (maximum 25). The same 25 items were administered as both pre-test and post-test, but item order was randomized by session. We acknowledge that reusing identical items can introduce a practice or testing effect that may inflate post-test performance independent of instruction.

E. Validity and Reliability Evidence

Content validity was established through expert review by three science specialists (Division Science Coordinator, Science Education Program Specialist, and SHS Science Coordinator). The mean content validity rating was 4.2/5, indicating good alignment with intended competencies.

A pilot test with a comparable Grade 11 cohort was administered twice, with a two-week interval between administrations. Internal consistency, estimated via KR-21 [19], was 0.81, indicating acceptable reliability for classroom research.

F. Fidelity of Implementation

To document fidelity: Flip application analytics logs captured whether each experimental-group student viewed the assigned micro lecture and posted a response before class. Teacher checklists verified that (i) a micro lecture video and prompt were uploaded by 6 PM daily, (ii) worksheets were posted, and (iii) representative student responses were addressed in the next lesson.

Attendance and participation records documented the presence and engagement in peer instruction and collaborative tasks for both groups. No major deviations from the planned procedures were recorded across the seven-day intervention window, which comprised two assessment days and five core instructional sessions.

G. Data Analysis

Given that the outcome variable was a bounded count score (0–25), we treated distributions as potentially non-normal and emphasized nonparametric analyses:

- **RQ1**: Experimental group pre–post gains were tested using the Wilcoxon signed-rank test.
- **RQ2**: Control group pre–post gains were tested using the Wilcoxon signed-rank test.
- **RQ3**: Between-group differences in post-test scores were tested using the Mann-Whitney U test.

For each test, we report effect sizes using $r = |Z| / \sqrt{N}$, where N is the number of paired observations (for Wilcoxon) or total pooled N (for Mann-Whitney).

As a robustness check, we also transformed each student's raw score to a percentage (0–100) and ran independent-samples t-tests. Prior to doing so, we inspected normality (Shapiro–Wilk) and homogeneity of variance (Levene's test) and found no statistically significant violations. We also compared pre-test means to confirm baseline equivalence and computed Cohen's d. Because students were individually randomized and the same instructor taught both groups, we did not apply cluster corrections.

H. Ethics

This study was reviewed and approved by the Department of Education, Division of San Carlos Research

Office—Research Ethics Committee. The participants were minors; written parental consent and written student assent were obtained. Participation was voluntary, and students could withdraw at any time without penalty. Only de-identified scores and activity logs were analyzed. Data were stored on password-protected, institutionally managed drives and will be retained for two years post-publication, consistent with the Philippine Data Privacy Act (RA 10173).

IV. RESULTS

As shown in Table 1, the Mann-Whitney U test revealed no statistically significant difference between groups at pre-test ($U = 483.00$, $p = 0.126$), with experimental (M rank = 39.20) and control (M rank = 31.80) groups demonstrating comparable baseline performance. This equivalence establishes that observed post-test differences can be attributed more confidently to the Flip application intervention rather than to pre-existing group differences.

Table 1. Baseline equivalence between groups (pre-test scores)

Group	n	Mean Rank	Sum of Ranks	U	Z	p-value
Experimental Group	35	39.2	1372	483	−1.529	0.126
Control Group	35	31.8	1113	-	-	-

Note: Mann-Whitney U = 483.00, Wilcoxon W = 1113.00. The non-significant result (p = 0.126) indicates no statistically significant difference between groups at pre-test.

Table 2. Experimental group pre–post comparison (Wilcoxon signed-rank test)

Variable	SD	W	Z	p-value	Effect Size (r)	Interpretation
Pre-test	3.02	-	-	-	-	-
Posttest	2.96	595	−5.093	<0.001	0.86	Large

Note. IQR = Interquartile Range (75th–25th percentile); Mdn = Median; W = Wilcoxon test statistic (sum of positive ranks). All 35 participants showed improvement or no change (34 improved, one tie). The enormous effect size (r = 0.86) indicates significant growth.

The Wilcoxon signed-rank test (Table 2) showed a statistically significant improvement in the experimental group's scores from the pre-test (Mdn = 10.00) to the post-test (Mdn = 24.00), with a large effect size ($r = 0.86$, $W = 595.00$, $Z = –5.09$, $p < 0.001$). This effect size greatly exceeds those reported in prior meta-analyses of flipped-classroom models in health professions education (SMD = 0.33, 95% CI [0.21, 0.46]) and in secondary/postsecondary settings ($d \approx 0.36$) [7, 11]. It also surpasses the moderate K-12 effect ($d^* = 0.54$) that was reported [20]. The large gain observed here may stem from the combination of structured pre-class videos, peer-mediated Flip application discussions, and built-in formative quizzes, all aligned with design principles to bolster the efficacy of the flipped-classroom approach [11, 17]. Moreover, the reduction in IQR (4.00→3.00) suggests elevated and more uniform mastery across participants, indicating that the intervention effectively closed skill gaps.

As presented in Table 3, the Wilcoxon signed-rank test revealed a statistically significant improvement in the control group's scores from pre-test (Mdn = 9.00, IQR = 6.00) to post-test (Mdn = 17.00, IQR = 5.00), with a considerable effect size ($W = 630.00$, $Z = −5.16$, $p < 0.001$, $r = 0.87$). All 35 participants (100%) demonstrated equal or higher post-test scores, with a median increase of 89% (from 9.00 to 17.00). The reduction in IQR (6.00 to 5.00) suggests marginally more consistent performance post-intervention. This large effect size is comparable to, and even slightly exceeds, typical gains under non-flipped formats reported in some meta-analyses, likely to reflect powerful testing effects when identical instruments are reused [20].

Table 3. Control group pre–post comparison (Wilcoxon signed-rank test)

Variable	n	M	SD	W	Z	p-value	Effect Size (r)	Interpretation
Pre-test	35	8.66	3.84	-	-	-	-	-
Posttest	35	17.34	3.38	630	−5.163	<0.001	0.87	Large

Note. IQR = Interquartile range (75th–25th percentile); Mdn = Median; W = Wilcoxon test statistic (sum of positive ranks). All 35 participants showed improvement or no change (35 improved). The large effect size (r = 0.87) indicates significant growth.

Table 4. Between-group post-test comparison (Mann–Whitney U test)

Group	n	Mean Rank	U	Z	p-value	Effect Size (r)	Interpretation
Experimental	35	49.14	135	−5.655	<.001	0.68	Large
Control	35	21.86	—	—	—	—	—

Note. IQR = Interquartile Range (75th–25th percentile); U = Mann-Whitney test statistic; Effect size calculated as rank-biserial correlation ($r = |Z|/\sqrt{N}$, where N = 70). The significant result ($p < 0.001$) with a medium effect size indicates higher post-test scores in the experimental group.

Table 4 presents the Mann-Whitney U test revealed significantly higher post-test scores in the experimental group (Mdn = 24.00, IQR = 7.00) compared to the control group (Mdn = 17.00, IQR = 5.00), with a large effect size ($U = 135.00$, $Z = −5.66$, $p < 0.001$, $r = 0.68$). This between-group difference complements the within-group improvements: while both groups showed near-perfect improvement rates (experimental: 97%, r = 0.86; control: 100%, r = 0.87), the experimental group's superior post-test performance—with a 140% median increase versus the control's 89%—demonstrates the intervention's added value. The significant between-group effect (r = 0.68) indicates substantial separation in the score distributions, likely due to the experimental intervention's efficacy and the control group's lower ceiling (max score = 22 vs. experimental = 25). Notably, the control group's marginally higher within-group effect size (r = 0.87 vs. 0.86) reflects more uniform improvements across all participants. In contrast, the experimental group's greater median gain highlights its ability to drive exceptional absolute growth despite minor individual variability. These findings underscore the intervention's dual strength: fostering consistent improvement while enabling higher peak performance. This effect exceeds the small-to-moderate between-group effects ($d \approx 0.19–0.50$) typically reported in flipped-classroom meta-analyses across disciplines [10–12]. The substantial separation in score distributions suggests that the Flipped Application intervention maintained consistency and enabled higher achievement ceilings, likely through enhanced student engagement, immediate peer feedback, and iterative self-assessment.

Although both groups exhibited large within-group effect sizes ($r \approx 0.86–0.87$), the experimental group's median percentage gain (140% vs. 89%) underscores its potency in driving absolute growth. The control group's marginally higher r reflects uniform moderate gains, whereas the Flipped Application approach fostered both broad and peak performance increases. This dual strength aligns with findings that well-designed flipped models can yield both depth and breadth of learning [19]. These findings support integrating Flipapplication-style video-discussion tools into senior high chemistry, suggesting that teacher training programs incorporate flipped classroom design workshops. Curriculum specialists should consider embedding asynchronous content delivery and structured peer interactions in DepEd's continued K-12 reforms to optimize conceptual mastery and critical thinking.

V. Discussion

This section interprets the findings in relation to the research questions, theoretical framing, and prior literature. We also temper the magnitudes of the observed effects by considering design features and potential confounds.

A. Why Did the Flipped Application Condition Outperform Traditional Instruction?

The Flipped Application model required students to (i) preview core ideas through concise micro lectures, (ii) articulate those ideas in their own words on the Flip application, and (iii) listen to or read peers' explanations before class. This cycle likely primed students' conceptual frameworks and revealed misconceptions early, allowing valuable class time to address specific gaps rather than reteach content broadly. This reflects the "prepare–engage–consolidate" logic of effective flipped instruction reported in prior work [5, 11, 12].

Peer video responses on the Flip application also functioned as low-stakes oral rehearsals of scientific language. Prior studies argue that requiring students to verbalize disciplinary reasoning supports both cognitive processing and communication skills, particularly in science disciplines, where students must justify claims with evidence [13–15]. In this study, students in the Flipped Application condition did not just consume information; they had to produce explanations, critique each other's reasoning, and then defend or revise their thinking in class. This aligns with a social constructivist view in which learning is strengthened when students externalize and negotiate meaning with peers.

In contrast, students in the traditional group largely encountered new content for the first time during class. Although they also showed strong within-group gains ($r = 0.87$), their opportunities for explanation, peer feedback, and iterative clarification were more limited, largely driven by teachers. This may explain why the Flipped Application group's post-test distribution shifted upward more dramatically.

B. Relation to Previous Research

The large within-group effect sizes ($r \approx 0.86$–0.87) and the large between-group effect ($r \approx 0.68$; Cohen's $d \approx 0.82$) exceed the small-to-moderate mean effects commonly reported in meta-analyses of flipped instruction (often around $d = 0.3$–0.5 in secondary and postsecondary STEM and K–12 samples [10–12, 20]. Therefore, at the same time, the results suggest that the Flipped Application model can accelerate mastery of fourth-quarter chemistry competencies in a Grade 11 STEM context in the Philippines. We emphasize that these effect sizes should not be generalized uncritically to longer courses, different topics, or different schools.

C. Engagement, Communication Skills, and STEM Readiness

Beyond test scores, the design of the intervention explicitly attempted to cultivate two skills emphasized in Philippine science reform efforts: (1) the ability to articulate and defend scientific reasoning orally, and (2) the ability to monitor one's own understanding and seek clarification. Although this study did not directly measure gains in oral communication skills, integrating the Flip application reflections with in-class peer instruction was intended to address documented weaknesses in scientific communication among Filipino learners [1, 4]. Future studies should supplement achievement tests with performance-based assessments of communication, such as recorded scientific explanations, peer feedback quality, or structured oral exams.

VI. Conclusion

This randomized pre-test–post-test study provides classroom-based experimental evidence that a Flipped Application instructional model, pre-class micro-lecture videos, structured Flip application peer reflection, and in-class collaborative problem-solving can produce greater short-term gains in Grade 11 chemistry achievement than traditional lecture-based instruction in a Philippine senior high school setting. Both the Flipped Application and traditional groups showed statistically significant improvements over the intervention. However, the Flipped Application group achieved substantially higher post-test performance and a large between-group effect, which we interpret as arising from early activation of prior knowledge, opportunities for peer explanation, and targeted in-class practice.

These results, however, were obtained from a seven-day sequence consisting of two assessment days and five core instructional sessions, which used the same 25-item test at pre- and post-assessment, and involved a single teacher in a single school, so the very large effect sizes and conclusions about impact should be treated with caution. Future research should therefore extend the Flipped Application model across longer time frames and multiple sites, employ parallel or partially novel assessment forms, and incorporate direct measures of engagement, oral scientific communication, and self-regulated learning to determine whether similar benefits can be sustained and generalized in broader Philippine STEM education contexts.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

N.T.H. conceived the study, coordinated implementation, and led data analysis and the research process. N.R.S.H. contributed to instrument development, supported data interpretation, and co-drafted and revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgment

The authors acknowledge the Department of

Education—Schools Division Office (SDO) of San Carlos, the principal and science faculty of Don Carlos Ledesma National High School, and the Grade 11 STEM students who participated in this research. We also acknowledge the Division Research Ethics Committee for approving the study protocol.

REFERENCES

[1] Department of Education. (2019). Program for International Student Assessment (PISA) 2018 Philippine National Report. [Online]. Available: https://www.deped.gov.ph/wp-content/uploads/2019/12/PISA-2018-Philippine-National-Report.pdf

[2] OECD. (2023). *PISA 2022 Results: Country Note for the Philippines*. OECD Publishing. [Online]. Available: https://doi.org/10.1787/philippines-2022-en

[3] I. V. S. Mullis, M. O. Martin, P. Foy, and M. Hooper, "TIMSS 2019 international results in science," *TIMSS & PIRLS International Study Center*, 2020.

[4] A. B. I. Bernardo, M. O. Cordel, M. O. Calleja, J. M. M. Teves, S. A. Yap, and U. C. Chua, "Profiling low-proficiency science students in the Philippines using machine learning," *Humanities and Social Sciences Communications*, vol. 10, 192, 2023. doi: 10.1057/s41599-023-01705-y

[5] S. J. DeLozier and M. G. Rhodes, "Flipped classrooms: A review of key ideas and recommendations for practice," *Educational Psychology Review*, vol. 29, no. 1, pp. 141–151, 2017. doi: 10.1007/s10648-015-9356-9

[6] K. S. Chen, L. Monrouxe, Y. H. Lu, C. C. Jenq, Y. J. Chang, Y. C. Chang, and P. Y. C. Chai, "Academic outcomes of flipped classroom learning: A meta-analysis," *Medical Education*, vol. 52, no. 9, pp. 910–924, 2018. doi: 10.1111/medu.13616

[7] K. F. Hew and C. K. Lo, "Flipped classroom improves student learning in health professions education: A meta-analysis," *BMC Medical Education*, vol. 18, 38, 2018. doi: 10.1186/s12909-018-1144-z

[8] J. F. Eichler and J. Peeples, "Flipped classroom modules for large enrollment general chemistry courses: A low-barrier approach to increase active learning and improve student grades," *Chemistry Education Research and Practice*, vol. 17, no. 2, pp. 197–208, 2016. doi: 10.1039/C5RP00159E

[9] A. B. Flynn, "Structure and evaluation of flipped chemistry courses: Organic & spectroscopy, large and small, first to third year, English and French," *Chemistry Education Research and Practice*, vol. 16, no. 1, 198–211, 2015. doi: 10.1039/C4RP00224E

[10] L. Cheng, A. D. Ritzhaupt, and P. Antonenko, "Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis," *Educational Technology Research and Development*, vol. 67, pp. 793–824, 2019.

[11] D. C. D. van Alten, C. Phielix, J. Janssen, and L. Kester, "Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis," *Educational Research Review*, vol. 28, pp. 1–18, 2019. doi: 10.1016/j.edurev.2019.05.003

[12] Z. Turan, "Evaluating whether flipped classrooms improve student learning in science education: A systematic review and meta-analysis," *Scandinavian Journal of Educational Research*, vol. 67, no. 1, pp. 1–19, 2023. doi: 10.1080/00313831.2021.1983868

[13] A. Kleftodimos and A. Triantafillidou, "Using the flip application video platform to practice science oral communication," *Tech. Trends*, vol. 67, no. 2, pp. 294–314, 2023. doi: 10.1007/s11528-022-00801-1

[14] L. Mercer-Mapstone and L. Kuchel, "Core skills for effective science communication: A teaching resource for undergraduate science education," *International Journal of Science Education, Part B*, vol. 7, no. 2, pp. 181–201, 2017. doi: 10.1080/21548455.2015.1113573

[15] N. N. Miskam and A. Saidalvi, "The use of the Flip application for teaching oral presentation skills to engineering students," *International Journal of Recent Technology and Engineering*, vol. 8(1C2), pp. 536–541, 2019.

[16] M. Rahman and J. E. Lewis, "Future of the flipped classroom in chemistry education," *Journal of Chemical Education*, vol. 98, no. 3, pp. 789–798, 2021. doi: 10.1021/acs.jchemed.1c01115

[17] D. C. D. V. Alten, C. Phielix, J. Janssen *et al.*, "Self-regulated learning support in flipped learning videos enhances learning outcomes," *Computers & Education*, vol. 158, 104000, 2020.

[18] N. D. Nayeri, F. A. Noodeh, H. S. Nia, A. Yaghoobzadeh, and A. H. Goudarzian, "Statistical procedures used in pre-test–post-test control group design: A review of papers in five Iranian journals," *Acta Medica Iranica*, vol. 61, no. 10, pp. 584–591, 2023. doi: 10.18502/acta.v61i10.15657

[19] G. F. Kuder and M. W. Richardson, "The theory of the estimation of test reliability," *Psychometrika*, vol. 2, no. 3, pp. 151–160, 1937.

[20] G. Zhu, "Is flipping effective? A meta-analysis of the effect of flipped instruction on K-12 students' academic achievement," *Educational Technology Research and Development*, vol. 69, pp. 733–752, 2021. doi: 10.1007/s11423-021-10028-1