Correlation between Happy Puzzle Hands-On Media, Student Motivation, and Writing Skills: Evidence from Vocational High School EFL Learners

Afredo Riski Ramadhan | English Language Education, Universitas Islam Majapahit, Mojokerto, Indonesia
Wiwik Mardiana | English Language Education, Universitas Islam Majapahit, Mojokerto, Indonesia
Ahmad Iklil Saifulloh | English Language Education, Universitas Islam Majapahit, Mojokerto, Indonesia
afredoriski64@gmail.com

Abstrak

Conditional sentences remain one of the most persistent sources of difficulty for EFL learners, and conventional grammar instruction rarely gives teachers a way to intervene before that difficulty hardens into error. This study tested whether pairing a flipped classroom model with learning analytics drawn from Google Classroom could raise senior high school students' mastery of conditional sentence writing beyond what conventional instruction produces. Using a quasi-experimental, unequal control-group design, two intact eleventh-grade classes at SMAN 1 Bangsal, Indonesia ($n = 36$ each) were assigned to an experimental group taught through a Learning Analytics-Supported Flipped Classroom (LASFC) model and a control group taught conventionally, across an eight-week intervention. Writing ability was measured through a pre-test and post-test scored on four dimensions: grammatical accuracy, contextual appropriateness, structural complexity, and unity in writing. Welch's t-test showed a statistically significant post-test advantage for the experimental group, $M = 83.33$ ($SD = 8.91$), over the control group, $M = 79.22$ ($SD = 3.32$), $t(44.554) = 2.595$, $p = .013$, with a medium effect size (Cohen's $d = 0.61$). A secondary analysis of normalized gain produced an apparently contradictory result: the experimental group's gain ($\langle g \rangle = 0.16$) was lower than the control group's ($\langle g \rangle = 0.53$). This study argues that the discrepancy is best explained by a ceiling effect tied to the experimental group's markedly higher pre-test score ($M = 80.22$ versus $M = 56.22$), not by any weakness in the intervention. The findings support the use of learning-analytics data—material-access records, quiz scores, and assignment-completion timestamps—as a diagnostic layer within flipped instruction, and they caution against relying on normalized gain alone when comparing groups with unequal starting points.

Kata Kunci: *learning analytics, flipped classroom, conditional sentences,* EFL *writing, ceiling effect*

Introduction

Conditional sentences sit at an awkward intersection in English grammar pedagogy: they are indispensable for hypothesizing, arguing, and reasoning about consequences in academic writing, yet their five distinct forms and shifting time reference make them one of the structures EFL learners misuse most often. Survey-based research places the scale of the problem in stark terms roughly three-quarters of EFL respondents in one study rated conditional clauses as difficult or very difficult to use (Zheng, 2022) and error-analysis work traces the difficulty to specific, recurring confusions, such as substituting the present progressive for the present simple in a zero-conditional if-clause (Harshalatha & Sreenivasulu, 2024). These are not isolated slips. They recur because the grammatical form and the communicative function of conditionals are learned separately in most classrooms, and rarely reconnected in a way that sticks.

Part of the reason lies in how conditional sentences are typically taught. Convention favors a lecture-then-practice sequence: the teacher explains the rule, then assigns exercises to be completed alone, with feedback arriving too late, if at all, to correct the specific misunderstanding a student had while writing (Harshalatha & Sreenivasulu, 2024). Because classroom time is scarce, teachers have little room to give individualized feedback on a structure as form-heavy as conditionals, and students are left practicing an error until it stabilizes as a habit. Broader research into ESL/EFL grammar error corroborates this pattern beyond conditionals specifically: prepositions, verb tense, and sentence structure are the recurring casualties of instruction that explains once and moves on (Chaudhary & Al Zahrani, 2020), and grammar instruction more generally is where EFL learners report the steepest difficulty, precisely because it is treated as a fixed body of rules to be transmitted rather than a skill to be practiced with feedback (Samad, 2022).

The flipped classroom model was designed to answer exactly this kind of time constraint. By moving direct instructionvideos, readings, and worked examples outside class hours, it frees class time for the practice and feedback that grammar instruction actually needs (Chen et al., 2025). The evidence for its effectiveness in language learning is no longer preliminary: a recent meta-analysis reports an aggregate effect size of $d = 0.68$ on English achievement (Chen et al., 2025), flipped instruction has been shown to outperform conventional teaching across a range of outcome measures (Qi, 2024), and in writing specifically, flipped designs have improved complexity, accuracy, and fluency simultaneously rather than trading one for another (Lin et al., 2022). For a structure like conditional sentences, where students need repeated, corrected practice more than repeated explanation, this reallocation of class time is the model's core appeal.

That appeal comes with a caveat, however. A flipped classroom without a mechanism for tracking what students actually did before class is a flipped classroom flying blind: a teacher cannot target feedback at problems they cannot see. Learning analytics closes that gap by turning the digital traces of pre-class study material-access frequency, time on task, quiz performance into information a teacher can act on (Mangaroska & Giannakos, 2019). Its documented uses in personalized learning span the full instructional cycle, from diagnosing where a learner stands, to grouping students by need, to predicting who is at risk, to visualizing engagement in real time (Khor, 2024), and feedback systems built on this kind of data have been linked to measurable gains in learning outcomes (Cuéllar & Contero, 2025). Combining the two models flipped delivery plus analytics-informed feedback has already shown promise for speaking skills (Shao & Liu, 2021) and for adaptive instruction more broadly (Yaseen et al., 2025), and the underlying logic is straightforward: analytics tells the teacher where the flip is failing, in time to intervene (Sharif & Atif, 2024).

Despite this convergence of evidence, three gaps remain visible once the literature is read specifically for conditional-sentence instruction. First, flipped-classroom research in EFL writing remains comparatively thin next to the volume of work on general language skills (Publikoa & Publikoa, 2017). Second, almost none of the flipped-classroom studies that do target EFL grammar instrument their intervention with learning analytics in any systematic way—engagement is assumed rather than measured. Third, Indonesian senior high schools are nearly absent from this literature, despite representing exactly the setting limited class time, curriculum pressure, uneven digital infrastructure where a data-informed flipped model would need to prove itself before wider adoption. This study was designed to sit at the intersection of those three gaps: it tests a Learning Analytics-Supported Flipped Classroom (LASFC) model, built specifically around conditional sentence writing, in an Indonesian public senior high school.

Three features distinguish this study from the work it builds on. First, rather than treating "writing" or "grammar" as an undifferentiated construct, it isolates one grammatical structure conditional sentences across all five forms (zero, first, second, third, and mixed) and measures improvement along four specific writing dimensions: grammatical accuracy, contextual appropriateness, structural complexity, and unity in writing. Second, it operationalizes learning analytics through the 4W Reference Model (Chatti et al., 2012), using authentic Google Classroom data material-access logs, quiz scores, and assignment-completion timestamps as the raw material for feedback, rather than treating analytics as a peripheral add-on. Third, it locates this intervention in a public Indonesian senior high school, a context where the infrastructure and support conditions for data-driven flipped learning are far from guaranteed, which makes the results more informative about real-world feasibility than a study run under ideal technological conditions would be.

Accordingly, this study asks a single guiding question: does an LASFC model produce significantly greater improvement in students' mastery of conditional sentence writing than conventional instruction, and if so, along which dimensions of writing mastery is that improvement concentrated? Answering this question is expected to contribute in two directions. Theoretically, it tests whether the learning-analytics cycle capture, report, predict, act, refine (Valenzuela et al., 2021) holds up when applied to a single, well-defined grammatical target rather than to writing or language ability in the abstract. Practically, it offers Indonesian English teachers a concrete, low-cost template for using data that a free LMS like Google Classroom already collects, without requiring specialized analytics software.

The significance of this contribution extends beyond a single school. Pedagogically, it supplies empirical evidence, from a context where such evidence is scarce, on whether combining learning analytics with flipped delivery is worth the extra design effort it demands relative to a flip run without any data layer. Technologically, it speaks to a broader question in Indonesian education about how far a widely available, free platform like Google Classroom can substitute for purpose-built analytics dashboards that most public schools cannot afford; if a teacher-managed spreadsheet of access logs and quiz scores can meaningfully improve outcomes, the barrier to adoption is organizational rather than financial. From an equity standpoint, personalized, data-informed feedback is one of the few instructional resources that scales without requiring more classroom hours a relevant consideration in a system where teacher time, not curriculum content, is often the binding constraint. The findings also bear on curriculum and teacher-development planning: if learning-analytics literacy measurably changes what a teacher can do with an already-flipped lesson, then training programs preparing English teachers for technology integration have a concrete skill to build toward, rather than a general exhortation to "use data" left unspecified. Finally, because conditional sentences are only one of several grammatical structures that Indonesian EFL learners persistently misuse, the methodological approach tested here narrowing the outcome to one structure, scored along four explicit dimensions offers a template that could be reapplied to other structures without redesigning the underlying model.

Literature Review

Five prior studies most directly shaped the design of this research. Kanwal (2024) found flipped instruction outperforming conventional teaching for grammar writing built around inflectional morphemes, but without any learning-analytics component. Ustun et al. (2023) tested a learning-analytics-based feedback intervention with higher-education computer-course students and found significant gains in achievement and self-directed learning, though in a setting far removed from EFL writing. Ariani et al. (2024) confirmed that a Flipped Digital Classroom improves general English writing skill, amplified by student engagement, again without analytics or a specific grammatical target. Öztürk and Çakıroğlu (2021) showed that flipped designs built on self-directed learning strategies improve EFL language skills broadly, and Chang (2023) demonstrated that even low-proficiency EFL learners benefit from flipped instruction across achievement, motivation, and self-efficacy. Table 1 summarizes how these five studies relate to the present one and the gap each leaves unaddressed. Read together, they reveal a clean division of labor in the literature: studies that use learning analytics rarely address EFL writing, and studies that address EFL writing through flipped instruction rarely use analytics; Indonesian senior high schools are essentially absent from either strand, which is the specific space this study occupies.

Learning analytics is defined, for instructional purposes, as the practice of collecting and interpreting data about learners in order to improve both learning and the environments in which it happens (Long & Siemens, 2011). The 4W Reference Model (Chatti et al., 2012) structures this practice around what data is collected, who is involved, why it is collected, and how it is analyzed; mapped onto this study, Google Classroom's material-access, quiz, and submission data constituted the What, the teacher and the 36 experimental-class students the Who, flagging at-risk students and calibrating feedback the Why, and weekly review of grade reports and Google Forms summaries the How. Connectivist theory (Siemens, 2005) reframes a student's digital footprint as a diagnostic signal of where understanding breaks down rather than a mere compliance record, a claim supported by the broader datafication-of-education literature (Williamson et al., 2020) and by evidence that interventions producing reliable gains combine fast feedback turnaround, actionable guidance, and a clear translation from data pattern to instructional decision (Ifenthaler & Yau, 2020; Sharif & Atif, 2024). These strands converge on the operational claim that raw LMS activity only becomes learning analytics once it is deliberately converted into an instructional decision, and that platform sophistication matters far less than the consistency with which collected data is reviewed and acted upon (Mangaroska & Giannakos, 2019) the principle that justifies treating an ordinary LMS like Google Classroom as sufficient infrastructure for this study.

The Flipped Learning Network's FLIP framework (2014) names four components a properly designed flipped classroom needs: a Flexible Environment that accommodates different paces and modes of learning; a shift in Learning Culture from teacher-delivered content to learner-constructed understanding; Intentional Content, chosen specifically because direct instruction is the least effective use of class time for it; and a Professional Educator who observes, gives feedback, and assesses in real time. Empirical support for the flexible-environment component is strong (Strelan et al., 2020, d = 0.40 across 114 studies), and Bond's (2020) review of K–12 implementations adds that engagement gains are most consistent when pre-class content is short, well-sequenced, and paired with a low-stakes comprehension check. Conditional-sentence formulas fit the intentional-content criterion well, since they can be explained once on video and referenced repeatedly, while judgment calls about which conditional type fits a given context benefit far more from face-to-face, in-the-moment feedback. The professional-educator component is where learning analytics earns its place inside the framework: teachers who received weekly participation and quiz reports delivered more precisely targeted support than teachers working from impression alone (Ustun et al., 2023).

No single theory fully accounts for why an LASFC model should outperform conventional instruction on a grammar-heavy writing task; the plausible explanation draws on constructivist learning theory, self-regulated learning theory, cognitive load theory, and personalized learning theory working together, consistent with Schuitema et al.'s (2022) argument that flipped-classroom effectiveness hinges on learner autonomy, teacher engagement, assessment design, and structured feedback combined rather than acting in isolation. Operationally, the study followed the five-stage learning-analytics cycle proposed by Valenzuela et al. (2021) Capture, Report, Predict, Act, and Refine translating Google Classroom's raw activity logs (material-access frequency and duration, quiz scores, and submission timing) into weekly grade reports and Forms summaries, used to flag at-risk students, translate predictions into differentiated feedback, and refine subsequent material based on post-session performance (Seng & Chuan, 2023; Cabi & Türkoğlu, 2025). On this basis, H1 predicts that students taught through the LASFC model will show significantly greater mastery of conditional sentence writing than students taught conventionally; H0 predicts no such difference.

Table 1

Summary of Previous Studies and the Gap Addressed by This Study

Author (Year)	Focus and Method	Key Finding	Gap This Study Addresses
Kanwal (2024)	Flipped vs. conventional learning on grammar writing (inflectional morphemes); quasi-experimental	Flipped group significantly outperformed conventional group	No learning analytics; general grammar focus, not conditional sentences or senior high school
Ustun et al. (2023)	LA-based visual feedback and recommendations in a flipped classroom; mixed-methods, pretest–posttest, higher education	Significant gains in achievement and self-directed learning in the LA-supported group	Conducted in a university computer course, not EFL writing; this study applies the same logic to senior high school EFL conditional-sentence writing
Ariani et al. (2024)	Flipped Digital Classroom and student engagement on general English writing; quantitative experimental	Significant positive effect on writing skills, strengthened by engagement	No learning analytics component and no specific grammatical target; this study narrows the construct to conditional sentences with LA-based feedback
Öztürk & Çakıroğlu (2021)	Self-directed learning strategies within a flipped design; experimental	Positive impact on overall EFL language skill development	Focused on general language skills without analytics; this study operationalizes LA data through the 4W model for a specific grammar target
Chang (2023)	Flipped learning for low-proficiency EFL students; quasi-experimental	Experimental group outperformed control on achievement, grammar/writing, motivation, and self-efficacy	No analytics and no specific sentence type as the material focus; this study combines both for conditional sentences in an Indonesian senior high school

Method

This study used a quantitative, quasi-experimental design, specifically an unequal control-group design (Creswell, 2012), in which two pre-existing eleventh-grade classes rather than randomly assigned individuals served as the experimental and control groups, since class composition at SMAN 1 Bangsal is fixed administratively before any research can begin. Both groups completed a pre-test before treatment and a post-test after an eight-week intervention period, with only the experimental group receiving the LASFC model; Table 2 summarizes the design.

Table 2

Quasi-Experimental Design Pattern

Group	Pre-Test	Treatment	Post-Test
Experimental	O_1	X	O_2

Control	O_3	-	O_4

The study was conducted at SMAN 1 Bangsal, a public senior high school in Mojokerto Regency, East Java, over the even semester of the 2025/2026 academic year, selected because its English teachers identified conditional sentences as a persistent point of difficulty and because its computer laboratory and internet access made an LMS-based intervention logistically feasible. From a population of 144 eleventh-grade students across four parallel classes, purposive sampling selected two classes with comparable prior English achievement: class XI-6 (n = 36) as the experimental group and class XI-7 (n = 36) as the control group, with equivalence between the two groups subsequently checked empirically through the pre-test rather than assumed from prior grades alone.

Writing mastery was measured through a two-part test administered as both pre-test and post-test, with equivalent difficulty but different prompts across the two administrations to control for practice effects: Task Type A (Sentence Completion, worth 40 of 100 points) and Task Type B (Error Correction, worth 60 of 100 points), both spanning all five conditional types. Both tasks were scored using an analytical rubric built around four aspects grammatical accuracy, contextual appropriateness, structural complexity, and unity in writing each assessed on a 100-point scale across four proficiency bands (Advanced, 85–100; Proficient, 70–84; Developing, 50–69; Beginner, 0–49); Table 3 presents the indicators anchoring the Advanced and Beginner bands for each aspect. Content validity was established through review by two experienced English-teaching lecturers, construct validity through a pilot administration analyzed with Pearson product-moment correlation, and reliability through inter-rater reliability via Cohen's Kappa (target ≥ .75) and internal consistency via Cronbach's Alpha (target ≥ .70), all confirmed before the instrument was finalized, with two calibrated raters scoring every response in the main study.

Table 3

Scoring Indicators for the Four Writing-Assessment Aspects

Aspect	What It Measures	Advanced (85–100)	Beginner (0–49)
Grammatical Accuracy	Verb-form accuracy in the if-clause and main clause for the intended conditional type; subject–verb agreement; punctuation	Fully accurate and consistent verb forms; no grammatical errors	Frequent, severe errors that obscure the conditional structure
Contextual Appropriateness	Whether the conditional type matches the intended meaning and situation, and whether the condition–result relationship is logical	Precise match between conditional type and intended meaning; logical, realistic relationship	Conditional type does not match intended meaning; relationship is illogical or absent
Structural Complexity	Variety of conditional types and clause structures used; avoidance of repetitive patterns	Varied, appropriately complex structures across conditional types	Minimal or no structural variation; heavy reliance on a single basic pattern
Unity in Writing	Logical coherence between the if-clause and main clause as a single idea	Coherent, logically connected idea with no contradiction	If-clause and main clause fail to form a unified, logical idea

The intervention ran over four meetings per class across eight weeks: a pre-test session, two combined treatment sessions, and a post-test session. In the experimental class, sessions opened with the teacher's review of Google Classroom access-duration and quiz-score data to identify students needing extra guidance, followed by collaborative writing practice and, in later meetings, differentiated grouping and peer feedback based on the analytics dashboard. The control class covered the identical five conditional forms across the same four meetings through direct instruction, written exercises, and class discussion, without an LMS, analytics-informed feedback, or a flipped structure holding content constant while varying only the pedagogical model. The learning-analytics component followed the five-stage Capture–Report–Predict–Act–Refine cycle: Google Classroom captured material-access, quiz/formative, and assignment-submission data continuously (Table 4), which was converted weekly into grade reports and Forms summaries, used to flag students with low access rates or below-average scores who tended to be the same students who later struggled most with the third and mixed conditionals translated into differentiated feedback and scaffolded exercises, and reviewed weekly to adjust subsequent material and pacing.

Table 4

Learning Analytics Data Captured via Google Classroom and Its Instructional Use

Data Type	What Was Recorded	How It Was Used
Material access	Number of students who opened the video/reading and their access duration	Identified students who had not, or had only partially, accessed material, prompting a brief re-explanation at the start of class
Quiz/formative results	Scores per conditional type from short post-material quizzes	Mapped mastery by topic, grouped students by proficiency, and shaped differentiated writing practice
Assignment submission	On-time/late/missing status and timestamp for worksheets, pre-test, and post-test	Monitored task discipline and kept the research schedule on track
Student progress	Cumulative completion rate across material access, quizzes, and assignments per meeting	Tracked overall trajectory and flagged students needing additional guidance or enrichment

All quantitative data were analyzed in SPSS version 21 through three sequential procedures: descriptive statistics (mean, standard deviation, minimum, maximum) for the pre-test and post-test distributions; a Shapiro–Wilk normality test to determine whether parametric analysis was justified (with a Mann–Whitney U test as the alternative had normality been violated); and Levene's test of equal variances prior to the group-comparison test, with Welch's t-test used in place of the standard independent-samples t-test where this assumption was violated. Effect size was calculated using Cohen's d (small = 0.2, medium = 0.5, large ≥ 0.8), and improvement within each group was additionally quantified using normalized gain, $\langle g \rangle$ = (post-test score − pre-test score) / (maximum score − pre-test score), interpreted as low (< 0.3), medium (0.3–0.7), or high (≥ 0.7).

Results and Discussion

Table 5 presents the pre-test and post-test descriptive statistics for both groups. The two groups did not begin from the same baseline: the experimental group's pre-test mean (80.22) was 24 points above the control group's (56.22), despite both classes having been selected for comparable prior achievement, and both groups' standard deviations shrank from pre-test to post-test, most sharply in the control group (13.92 to 3.32). This pre-test gap is not incidental; it becomes the central interpretive problem resolved later in this chapter.

Table 5

Descriptive Statistics of Pre-Test and Post-Test Scores

Data	Group	n	M	SD	Min	Max
Pre-test	Experimental (XI-6)	36	80.22	12.47	52	100
Pre-test	Control (XI-7)	36	56.22	13.92	32	88
Post-test	Experimental (XI-6)	36	83.33	8.91	66	100
Post-test	Control (XI-7)	36	79.22	3.32	74	88

Shapiro–Wilk results (significance values of .224, .213, .550, and .070 for the four distributions, all above the .05 threshold) confirmed that each distribution was consistent with normality, justifying a parametric group-comparison test, while Levene's test indicated a violation of the equal-variances assumption between the two post-test distributions, $F = 34.046$, $p < .001$—unsurprising given how much tighter the control group's post-test spread ($SD = 3.32$) was than the experimental group's ($SD = 8.91$). Because this violation meant the "equal variances not assumed" row—Welch's t-test—had to be used to test H0 and H1, the analysis produced $t(44.554) = 2.595$, $p = .013$, a significant result at the .05 level leading to rejection of H0: the experimental group's post-test mean (83.33) exceeded the control group's (79.22) by 4.11 points, 95% CI [0.92, 7.30], an interval excluding zero and therefore supporting the difference as reliable.

Cohen's d, calculated from the pooled post-test standard deviation (6.72), was 0.61 a medium effect by conventional benchmarks, comparable to, if not somewhat larger than, the $d = 0.40$ that Strelan et al. (2020) report as the overall effect of flipped instruction across 114 studies. Table 6 presents the normalized-gain analysis that complicates this otherwise straightforward result: taken at face value it appears to contradict the t-test finding, since the group with the significantly higher post-test score shows the smaller normalized gain. This is resolved once the arithmetic of normalized gain is examined: because ⟨g⟩ is scaled by the distance remaining to the maximum possible score, a group starting near the ceiling has structurally less room to register a large gain, independent of how effective the intervention actually was. The experimental group's pre-test mean left only 19.78 points of headroom before the 100-point ceiling, versus 43.78 points for the control group more than double so the two ⟨g⟩ values are not measuring the same thing.

Table 6

Normalized Gain of Pre-Test and Post-Test Scores

Group	Pre-Test M	Post-Test M	⟨g⟩	Category
Experimental (XI-6)	80.22	83.33	0.16	Low
Control (XI-7)	56.22	79.22	0.53	Medium

Beyond the numerical results, the pattern of learning-analytics data collected over the eight weeks offered a diagnostic layer the post-test alone could not provide: students who logged low video-access rates or below-average quiz scores during the pre-class phase were disproportionately the same students who later produced the weakest post-test responses on the third and mixed conditionals, the two forms requiring the most complex verb-tense coordination. In Meeting 3, where engagement data diverged most sharply between students, the teacher used the analytics dashboard to split the class into differentiated working groups the specific mechanism this study argues underlies the LASFC model's advantage: not flipped delivery alone reallocating class time toward practice, but analytics data making that reallocated time selectively targeted rather than uniformly distributed.

This central finding is broadly consistent with, rather than an outlier within, the existing literature: Kanwal (2024) found flipped instruction outperforming conventional teaching for grammar writing generally, Ustun et al. (2023) found learning-analytics-based feedback improving both achievement and self-directed learning in a flipped setting, Ariani et al. (2024) and Öztürk and Çakıroğlu (2021) each found flipped designs improving English writing and language skill respectively, and Chang (2023) found the flipped advantage holding even for lower-proficiency learners. What this study adds is a narrower and more instrumented test of the same underlying claim, uniquely combining flipped delivery with systematic analytics-based feedback around a single, narrowly defined grammatical target in an Indonesian public senior high school a convergence test rather than a discovery of a wholly new effect. Cohen's d = 0.61 indicates the effect is large enough to matter pedagogically, somewhat above Strelan et al.'s (2020) general flipped-instruction baseline, a gap consistent with, though not decisive proof of, the theoretical claim that analytics-informed feedback improves on an unmonitored flip by letting a teacher target the specific gaps a class actually has.

The apparent contradiction between the significant post-test advantage and the lower normalized gain deserves particular care, since handled carelessly it could be misread as evidence against the intervention's effectiveness. The explanation is a ceiling effect rooted in the experimental group's unexpectedly high pre-test baseline (80.22, close to the 100-point ceiling) itself a reminder that general English proficiency and mastery of one specific grammatical structure are not interchangeable measures, and that purposive sampling based on overall grades cannot guarantee equivalence on a narrow construct measured for the first time at pre-test. This interpretation gains independent support from Heikkinen et al. (2022), whose systematic review found that only 46% of learning-analytics interventions aimed at supporting self-regulated learning showed a directly measurable positive effect, a finding attributed in part to exactly this kind of measurement artifact. The methodological lesson is that normalized gain should be reported alongside, never instead of, a direct between-groups comparison, and interpreted with explicit reference to each group's proximity to the measurement ceiling rather than compared as though the two groups started from the same position.

These results lend empirical weight to the theoretical architecture reviewed in Chapter 2: the 4W Reference Model's parameters functioned as intended, connectivist theory's claim that digital footprints carry diagnostic meaning was borne out by the alignment between access data and eventual performance on specific conditional types, and the five-stage Capture ReportPredict Act Refine cycle functioned as a genuine feedback loop rather than a one-time diagnostic. Practically, the analytics layer required no specialized software; Google Classroom's built-in grade reports and Forms summaries were sufficient to run the full cycle, which matters directly for Indonesian public schools where infrastructure and budget constraints often make analytics platforms feel out of reach; the third and mixed conditionals emerge as the forms where analytics-informed scaffolding earns its keep most clearly. For administrators, infrastructure adequacy is a precondition rather than a background detail, since the entire analytics cycle depends on reliable internet access and device availability. For teacher preparation, the operative skill is not tool proficiency but the interpretive habit of turning a grade-report number into a concrete classroom decision before the next meeting (Ahmad Iklil Saifulloh, 2015; Manthofani et al., 2024; Saifulloh, Retnaningdyah, & Mustofa, 2025c, 2025b, 2025a; Saifulloh, Retnaningdyah, Mustofa, et al., 2025; Saifulloh, 2019, 2021, 2024, 2026a, 2026b; Saifulloh et al., 2023; Saifulloh & Anam, 2022; Saifulloh & Irfan, 2021).

Several limitations qualify how far these findings can be generalized. The sample was confined to two intact classes at a single school, which limits statistical power and constrains generalization to other schools, grade levels, or grammatical structures. The eight-week intervention window cannot speak to whether the observed gains persist over time or transfer to conditional-sentence use outside a test context. The pre-existing gap between the two groups' baseline knowledge of conditional sentences, despite comparable general English grades, complicates direct comparison and is the direct source of the normalized-gain interpretation problem; a design with a true baseline-equivalence check on the specific construct being measured would have avoided this complication. Finally, the model's dependence on stable internet access and device availability means its results should not be assumed to generalize to schools with weaker technological infrastructure than SMAN 1 Bangsal's.

Conclusion

This study set out to test whether a Learning Analytics-Supported Flipped Classroom model improves EFL students' mastery of conditional sentence writing more than conventional instruction, using data drawn directly from Google Classroom rather than from a specialized analytics platform. The answer, on the primary measure, is yes: the experimental group significantly outperformed the control group on the post-test, $t(44.554) = 2.595$, $p = .013$, with a medium effect size ($d = 0.61$) that compares favorably with the broader flipped-classroom literature. The study also surfaced a result that initially looks like a contradiction the experimental group's lower normalized gain and traced that result to a ceiling effect produced by the experimental group's unexpectedly high pre-test baseline, rather than to any weakness in the intervention itself. That resolution carries a methodological lesson beyond this study's own findings: normalized gain is not a safe substitute for a direct between-groups comparison when the two groups do not start from comparable positions, and researchers working with unequal-baseline designs should report and interpret both statistics together rather than treating either in isolation.

Substantively, the findings support treating learning-analytics data material-access logs, quiz scores, and submission timestamps, all collected through a free LMS already available to most schools as a genuine diagnostic layer within flipped instruction, particularly for identifying which students are struggling with which specific conditional forms early enough to intervene. For English teachers, this offers a low-cost, infrastructure-light template for combining flipped delivery with data-informed feedback. For future research, three directions follow most directly from this study's limitations: replication with a larger and more heterogeneous sample to strengthen generalizability; a longer intervention window capable of testing retention and transfer beyond the immediate post-test; and explicit baseline-equivalence testing on the specific construct under study, not merely on general achievement, to avoid the ceiling-effect complication documented here. Extending the model to other grammatically demanding structures beyond conditional sentences would also help establish whether the pattern observed here significant post-test gains alongside a baseline-dependent normalized-gain result is specific to conditionals or a more general feature of analytics-supported flipped instruction.

Taken together, the value of this study lies less in demonstrating that flipped classrooms or learning analytics work in isolation both claims already have substantial support—than in showing what happens when they are combined narrowly, instrumented carefully, and tested in a setting where neither infrastructure nor student baseline can be taken for granted. The ceiling-effect finding, in particular, is offered here not as a caveat to be explained away but as a substantive methodological contribution in its own right: any researcher or practitioner comparing groups with unequal starting points on a bounded scale should expect normalized gain and a direct between-groups test to diverge, and should be prepared to interpret that divergence rather than treat it as a threat to the study's validity. If this study succeeds in anything beyond its immediate result, it is in making that interpretive move explicit enough that future work on learning-analytics-supported instruction, in EFL classrooms or elsewhere, need not rediscover it independently.

References

Ariani, F., Rahman, F., & Sultan, S. (2024). The effect of flipped digital classroom and student engagement on English writing skills. Studies in English Language and Education, 11(2), 976–997.

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. Computers & Education, 151, Article 103819. https://doi.org/10.1016/j.compedu.2020.103819

Cabi, E., & Türkoğlu, S. (2025). The impact of a learning analytics based feedback system on students' academic achievement and self-regulated learning in a flipped classroom. Interactive Technology and Education, 26(1).

Chang, D. Y. (2023). Flipping EFL low-proficiency students' learning: An empirical study. Language Teaching Research. https://doi.org/10.1177/13621688231165474

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. International Journal of Technology Enhanced Learning, 4(5–6), 318–331.

Chaudhary, A., & Al Zahrani, S. (2020). Error analysis in the written compositions of EFL students: A classroom study. International Journal of English Linguistics, 10(2), 357–366. https://doi.org/10.5539/ijel.v10n2p357

Chen, H., Chen, C., & Wu, W. V. (2025). Effects of flipped learning on language learning outcomes: A meta-analysis investigating moderators. SAGE Open, 15(2). https://doi.org/10.1177/21582440251331298

Creswell, J. W. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.). Pearson.

Cuéllar, Ó., & Contero, M. (2025). Personalized and timely feedback in online education: Enhancing learning with deep learning and large language models. Computers and Education: Artificial Intelligence.

Harshalatha, S., & Sreenivasulu, Y. (2024). Exploring academic writing needs and challenges experienced by ESL learners: A literature review. World Journal of English Language, 14(3), 406–412. https://doi.org/10.5430/wjel.v14n3p406

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2022). Supporting self-regulated learning with learning analytics interventions—A systematic literature review. Education and Information Technologies, 28(3), 3059–3088. https://doi.org/10.1007/s10639-022-11281-4

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. Educational Technology Research and Development, 68(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Kanwal, A. (2024). Examining the effect of flipped learning on proficiency in English writing grammar via inflectional morphemes. Theory and Practice in Language Studies, 14(4), 1118–1126.

Khor, E. T. (2024). A systematic review of the role of learning analytics in supporting personalized learning. Education Sciences.

Lin, V., Liu, G.-Z., & Chen, N.-S. (2022). The effects of an augmented-reality ubiquitous writing application: A comparative pilot project for enhancing EFL writing instruction. Computer Assisted Language Learning, 35(5–6), 989–1030. https://doi.org/10.1080/09588221.2020.1770291

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 46(5), 30–32.

Mangaroska, K., & Giannakos, M. (2019). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. IEEE Transactions on Learning Technologies, 12(4), 516–534. https://doi.org/10.1109/TLT.2018.2868673

Öztürk, M., & Çakıroğlu, Ü. (2021). Flipped learning design in EFL classrooms: Implementing self-regulated learning strategies to develop language skills. Smart Learning Environments, 8, Article 24.

Publikoa, N. U., & Publikoa, N. U. (2017). Flipping the EFL classroom in a secondary education setting: Students' perceptions and academic performance. ELT Journal Studies, 35–57.

Qi, P. (2024). A systematic review of flipped classroom approaches in language learning. Language Learning & Technology, 16(4).

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery, 10(3), Article e1355. https://doi.org/10.1002/widm.1355

Ahmad Iklil Saifulloh, S. I. (2015). *FILSAFAT KESEHARIAN Praktik Pendidikan, Bahasa, dan Sastra*.

Manthofani, M. T., Saifulloh, A. I., & Lindawati, R. (2024). *Transforming EFL Students ' Narrative Writing Through Gamification*. *2*(4), 221–228.

Saifulloh, A. I. (2019). The Implementation of Ecocriticism Approach for Healthy Earth and strengthen the English Writing Skill of Literary Appreciation Class. *IJOELL*, *01*(02), 10–17.

Saifulloh, A. I. (2021). Ecocriticsm as the Development of Teaching Materials in Literary Theories and Appreciation Courses. *International Journal of English Education and Linguistics (IJoEEL)*, *3*(1), 1–9. https://doi.org/10.33650/ijoeel.v3i1.1752

Saifulloh, A. I. (2024). *Kritik Sastra Lingkungan: Pengantar, Pendekatan, dan Implementasinya di dunia pendidikan* (Issue March). Tema Litera. https://www.researchgate.net/publication/376754428_Kritik_Sastra_Lingkungan_Penganta r_Pendekatan_dan_Implementadinya_di_dunia_pendidikan

Saifulloh, A. I. (2026a). *IMPLEMENTING MULTILITERACIES PEDAGOGY TO PROMOTE ENGLISH MULTIMODAL AND ECOLOGICAL LITERACY*.

Saifulloh, A. I. (2026b). *Introduction to Literature : Reading , Interpreting , and Analyzing Literary Texts*.

Saifulloh, A. I., & Anam, S. (2022). The Effectiveness of Pedagogical Ecocriticism to Enhance Environmental Awareness in Post Pandemic Era. *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, *14*(1), 1–18. https://doi.org/10.37680/qalamuna.v14i1.1130

Saifulloh, A. I., & Irfan, S. (2021). Implementasi Gerakan Literasi Digital Di Era Berkelimpahan Informasi Bagi Pendidikan Tinggi. *MATAPENA: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, *4*(2), 182–192. http://ejurnal.unim.ac.id/index.php/matapena/article/view/1287

Saifulloh, A. I., Retnaningdyah, P., & Mustofa, A. (2025a). Immersive Situated Practice Within a Local Curriculum for Hybrid Learning. *Veredas Do Diretio*, 2–30. https://doi.org/https://doi.org/10.18623/rvd.v22.n5.3587

Saifulloh, A. I., Retnaningdyah, P., & Mustofa, A. (2025b). *Literacy as Resistance : Critical Framing for Promoting Ecological Awareness*. 996–1012. https://doi.org/10.19105/ghancaran.vi.21731

Saifulloh, A. I., Retnaningdyah, P., & Mustofa, A. (2025c). *Theoretical Praxis in Environmental Education within EFL Contexts*. *1*(1), 1–14.

Saifulloh, A. I., Retnaningdyah, P., Mustofa, A., & Suhartono, S. (2025). Immersive Situated Practice Within a Local Curriculum for Hybrid Learning. *Veredas Do Direito*, *22*(5), e223587. https://doi.org/10.18623/rvd.v22.n5.3587

Saifulloh, A. I., Retnaningdyah, P., & Mustopa, A. (2023). *Damaging Discourse in Online News Trends Over Ecological Literacy*. *5*(4), 885–900. https://www.elitejournal.org/index.php/ELITE

Schuitema, J., Bakker, A., de Boer, H., & Poortman, C. (2022). Flipped classroom effectiveness and the factors that determine it. Educational Studies.

Seng, E., & Chuan, N. C. (2023). Integrating learning analytics into the flipped classroom: A study on data-enabled flipped learning. Open Access Library Journal, 10. https://doi.org/10.4236/oalib.1110692

Shao, M., & Liu, X. (2021). Impact of the flipped classroom on students' learning performance via meta-analysis. Journal of Sustainable Development, 9(9), 82–109. https://doi.org/10.4236/jss.2021.99007

Sharif, H., & Atif, A. (2024). The evolving classroom: How learning analytics is shaping the future of education and feedback mechanisms. Education Sciences.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2(1), 3–10.

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, Article 100314. https://doi.org/10.1016/j.edurev.2020.100314

Ustun, A. B., Zhang, K., Karaoğlan-Yilmaz, F. G., & Yilmaz, R. (2023). Learning analytics based feedback and recommendations in flipped classrooms: An experimental study in higher education. Journal of Research on Technology in Education, 55(5), 841–857. https://doi.org/10.1080/15391523.2022.2040401

Valenzuela, C. G., González, C. G., Rojas, A., & Tagle, M. (2021). Learning analytics in higher education: A preponderance of analytics but very little learning? International Journal of Educational Technology in Higher Education, 18, Article 15. https://doi.org/10.1186/s41239-021-00258-x

Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives. Teaching in Higher Education, 25(4), 351–365.

Yaseen, H., Mohammad, A. S., Ashal, N., Abusaimeh, H., & Ali, A. (2025). The impact of adaptive learning technologies, personalized feedback, and interactive AI tools on student engagement: The moderating role of digital literacy. International Journal of Data and Network Science.

Zheng, X. (2022). English as an international language project-based learning: A focus on young Chinese learners' beliefs. English Language Teaching, 15(8), 1–12. https://doi.org/10.5539/elt.v15n8p1