Jurnal Pustaka Indonesia (JPI)
2026, Vol. 6, No. 2, 342–359
https://doi.org/10.62159/jpi.v6i2.2588

OPEN ACCESS Research Article

# Effects of the Flipped Classroom on Conceptual Understanding and Learning Readiness in Vocational Mathematics: A Quasi-Experimental Study

Rahmi Mulia Yanti[1*], Ulfia Rahmi[2], Alwen Bentri[3], and Hani Fannisa[4]
[1-4] Universitas Negeri Padang, Padang, Indonesia

**Abstract**
Mathematics teaching in Indonesian vocational secondary schools (SMK) remains dominated by teacher-centred routines that leave many students with shallow conceptual understanding and weak learning readiness. This study compared a flipped-classroom package with conventional instruction in a quasi-experimental pre-test-post-test control-group design: two intact Grade XI classes at SMK Negeri 1 Painan were allocated to flipped ($n$ = 30) or conventional ($n$ = 31) conditions during a unit on circles and transformations. Because allocation was at class level and not random, class-specific influences are confounded with condition, and estimates are adjusted group differences, not causal effects. Conceptual understanding was measured with a double-marked rubric-scored essay test and learning readiness with a Likert questionnaire covering physical, psychological, cognitive, and environmental dimensions. A MANCOVA with both pre-tests as covariates returned a significant multivariate difference, Wilks' $\Lambda$ = .49, $F(2, 56)$ = 29.13, $p$ < .001, partial $\eta^2$ = .51. Follow-up ANCOVAs showed adjusted differences of 9.75 points on conceptual understanding, 95% CI [6.09, 13.40], partial $\eta^2$ = .33, and 8.24 points on learning readiness, 95% CI [5.32, 11.16], partial $\eta^2$ = .36, both surviving Holm correction. Normalized gains were medium in the flipped class and low in the conventional class, and estimates held under Welch and rank-based tests. With one class per condition, one teacher, and one school, these results describe a difference in this setting and cannot separate the model from novelty, access, and compliance.

**ARTICLE HISTORY**
Received : 16 April 2026
Revised : 26 June 2026
Accepted : 25 July 2026

**KEYWORDS**
Flipped classroom; learning readiness; vocational mathematics; conceptual understanding



**CORRESPONDING AUTHOR**
***Rahmi Mulia Yanti**, Universitas Negeri Padang, Padang, Indonesia. Email: 1985.yanti87antich@gmail.com

## Introduction

Mathematics occupies a strategic place in vocational secondary education because it underpins the technical reasoning, measurement, and data-handling that graduates carry into the workplace (Lubis & Siregar, 2024; Selvia & Arnawa, 2025). That general statement, however, is too loose to justify an intervention study, because what mathematics is for differs sharply by programme, and a claim of vocational relevance has to be made against the actual content taught and the actual occupation the students are training for. The participants here were students of the Office Management programme (Manajemen Perkantoran), and the instructional unit covered circles and geometric transformations. Table 1 sets out how each of the seven understanding indicators assessed in this study maps onto tasks in that programme, and it is worth stating plainly that the fit is closer for some indicators than for others.

**Table 1.** Mapping of the seven understanding indicators onto tasks in the vocational programme

| Understanding indicator | Task in the unit on circles and transformations | Corresponding task in the Office Management programme | Strength of the link |
|---|---|---|---|
| Interpreting | Convert between a transformation described in words, in coordinates, and in a diagram | Reading a scaled office layout plan and describing the arrangement it specifies | Direct |
| Exemplifying | Give an instance of a dilation with a stated scale factor | Producing a floor plan or seating diagram at a stated scale | Direct |
| Classifying | Decide whether a mapping is a translation, reflection, rotation, or dilation | Distinguishing a resized document layout from a repositioned or rotated one | Moderate |
| Summarising | State in one proposition what a composition of transformations does | Summarising the change made to a workflow diagram between two versions | Moderate |
| Inferring | Detect the pattern in a sequence of transformed figures | Detecting a trend in a series of charted monthly office data | Moderate |
| Comparing | Contrast the effect of a rotation with that of a reflection | Comparing two proposed room layouts on the same floor area | Direct |
| Explaining | Build a cause-and-effect account of why a circle's area scales with the square of the radius | Explaining why a proportional change in a chart dimension misrepresents a quantity | Moderate |

*Note.* The mapping is offered as the justification for teaching this unit in this programme and should be read as a claim about transfer of reasoning rather than about direct occupational application.

Read against Table 1, the honest position is intermediate. Translation, dilation, and rotation have direct counterparts in the scaled office layout plans, seating and workflow diagrams, and document and slide layout work that this programme requires, and circle geometry appears in the measurement of circular tables and workstations and in the interpretation of pie charts. What the unit does not do is train an occupational skill directly, in the way a drafting or construction programme would use the same content. The case for the unit rests on transfer of the reasoning rather than on the immediate applicability of the formulas, and the earlier framing of this manuscript, which invoked drafting, costing, and quality control in general, overstated a link that the programme in question does not support. That correction matters for interpretation, because the conceptual gains reported later should be read as gains in mathematical reasoning within a vocationally contextualised unit, not as evidence of improved occupational competence (Jitendra et al., 2016; Majid et al., 2025).

In practice, mathematics is too often experienced as a collection of formulas to be memorised rather than a body of ideas to be understood. National and international assessments have repeatedly placed Indonesian students at the lower end of the mathematics-reasoning distribution (Ningsih et al., 2025; Purnomo, 2017; Rahmah et al., 2025). The same pattern is visible at the classroom level, where students can reproduce a procedure but struggle to explain why it works or to transfer it to a problem they have not met before. A performance gap of this kind signals a difficulty with understanding rather than with memory, and it is precisely understanding that vocational work demands (Abadiyah et al., 2025; Nouwen et al., 2022).

The situation at the study site was established by diagnostic evidence rather than by teacher report alone, and the earlier version of this manuscript rested too heavily on the latter. Three sources were used. First, the preceding term's mathematics records for all Grade XI Office Management classes were examined against the school's minimum mastery criterion; across the two classes eventually studied, 47 of 61 students (77.0%). Second, the pre-tests administered in this study serve as a diagnostic baseline in their own right: mean conceptual understanding stood at about 45 points on a 0 to 100 scale in both classes, and an item-level breakdown showed that performance fell away sharply on the indicators that require explanation and inference rather than execution. Third, a short open-response survey administered before the unit asked students what made mathematics difficult and what they did before a lesson. Taken together these sources locate the problem more precisely than teacher impression: students arrive without prior contact with the material and can execute procedures without being able to justify them.

Conceptual understanding refers to the capacity to interpret, exemplify, classify, and explain mathematical ideas rather than merely execute steps (Ifrianti et al., 2019; Muktiadji et al., 2026). A student who understands a concept can move flexibly between its representations, recognise it in an unfamiliar guise, and justify the reasoning that links one idea to the next. In a vocational setting this flexibility matters, because the problems graduates meet on the job rarely arrive in the tidy form of a textbook exercise (Candra et al., 2022; Daud et al., 2024).

Learning readiness needs a more careful theoretical treatment than it received in the earlier version of this manuscript, because the way the construct is understood determines whether it can legitimately be treated as an outcome at all. Three readings are possible. On the first, readiness is a stable antecedent, a trait-like property of the learner that instruction inherits rather than alters; on that reading it belongs on the right-hand side of a model as a moderator or a covariate, and measuring it as an outcome would be a category error. On the second, readiness is a purely situational state, fluctuating with sleep, mood, and the events of a given morning; on that reading it is measurable but not a meaningful target, because a six-meeting intervention cannot be expected to move a quantity that varies day to day. On the third, readiness is intervention-responsive: partly dispositional and partly produced by the structure of instruction itself, and therefore stable enough to measure and malleable enough to change.

This study adopts the third reading, and the adoption carries commitments that are stated here so that the results can be judged against them. The four dimensions conventionally used to frame readiness, namely physical, psychological, cognitive, and environmental (Chikeme et al., 2024; Jamaris et al., 2021), are not equally malleable, and the flipped design does not act on them equally. Cognitive readiness, understood as possession of the prerequisite material at the moment the lesson begins, is the dimension the design targets most directly, because a pre-class task makes prior contact with the content a condition of participation rather than an act of unusual diligence. Psychological readiness, in the sense of confidence that one can follow the lesson, should follow from cognitive readiness by way of self-regulatory experience, since a student who has already met the material has grounds for expecting to cope (Sartori et al., 2013). Physical and environmental readiness, by contrast, concern rest, materials, home conditions, and social relations, and there is no plausible route by which six meetings of a rearranged lesson sequence could alter them; if they move, the more economical explanation is response bias rather than instruction.

Two consequences follow, and both shape the analysis reported below. First, readiness is modelled as an outcome, but a directional prediction is attached to it: the cognitive and psychological dimensions should move and the physical and environmental dimensions should not, and the dimension-level results are reported so that this prediction can be checked rather than assumed. Second, because readiness was measured by self-report after an intervention that students may have found more engaging than their usual lessons, a total score that rises uniformly across all four dimensions would be more consistent with expectancy and social-desirability effects than with the mechanism proposed here (Hikmat, 2022). The dimension profile therefore functions as an internal check on the readiness result, and it is treated as such in the discussion.

Both problems are sustained by the conventional, teacher-centred format that still prevails in many classrooms. Lessons proceed largely as one-way transmission followed by routine drill, which casts students as passive recipients and leaves little class time for exploration or discussion (Akmal et al., 2025; Kholifah et al., 2023; Raman et al., 2019). Just as importantly, the conventional format gives students no structured reason to engage with the material before the lesson, so they enter the room unprepared. The limited contact time available in SMK, already compressed by a heavy vocational workload, is then consumed by content delivery rather than by sense-making (Fajra et al., 2020; Maryanti & Nandiyanto, 2021).

The flipped classroom offers a direct response to this configuration. By moving first exposure to content outside the classroom through short videos and digital modules, and by reserving face-to-face time for discussion, problem solving, and application, the model is designed to deepen understanding and, at the same time, to require students to prepare in advance (Egara et al., 2025; Hafidzah et al., 2025; Tsai et al., 2017). The rearrangement changes not the content of a lesson but the place where each kind of cognitive work is done.

The empirical record is favourable but more modest than a selective reading suggests. A meta-analysis of mathematics outcomes reports an advantage for flipped designs (Egara et al., 2025), and a vocational study found gains in mathematical connection alongside reduced anxiety (Alsalhi, 2026), while investigations in general-secondary settings have documented improvements in conceptual understanding (Bhagat et al., 2016; Hamdan et al., 2022). Broader syntheses that pool across disciplines, however, place the average effect on learning in the small-to-moderate range and note substantial heterogeneity, with satisfaction sometimes unchanged or lower than in conventional teaching (Bada & S. Olusegun, 2016; Qolamani, 2023). Two moderators recur: the benefit is larger where in-class time is genuinely restructured rather than merely supplemented, and it depends on students actually completing the pre-class work, which many do not. Effects of the size reported below should therefore be treated as unusual for this literature rather than as typical of it, and the compliance data reported in the Method are part of what any reader needs in order to judge them.

Three gaps remain. First, most Indonesian flipped-classroom research has been conducted in general lower- and upper-secondary schools, whose students differ from SMK students in workload and vocational orientation (Herlambang, 2024; Muktiadji et al., 2026). Second, vocational studies have tended to emphasise problem solving and reasoning, leaving conceptual understanding comparatively unexamined (Lai et al., 2020; Mansour & Wardat, 2025). Third, learning readiness is seldom measured directly as an outcome, so the claim that flipping cultivates preparation is usually inferred rather than tested.

This study asks whether, after adjustment for prior conceptual understanding and prior learning readiness, Grade XI vocational students taught with a flipped-classroom package differ from those taught conventionally on the combined outcomes of conceptual understanding and learning readiness. The question is deliberately phrased as a comparison of adjusted group means rather than as an estimate of the effect of the flipped model, because with one intact class per condition the model is fully confounded with everything else that distinguishes the two classes. The secondary question asks whether the readiness gain, if present, follows the dimension profile predicted above, since a uniform rise across all four dimensions would point to response bias rather than to the proposed mechanism.

# Method

## *Research design*

The study employed a quasi-experimental pre-test-post-test control-group design. Random assignment of individual students was not feasible because classes at the school are already formed, so two intact classes were used (Sugiyono, 2016), with the experimental class receiving the flipped-classroom package and the control class conventional instruction over the same unit and the same time span. The design is summarised in Table 2. The inferential limits of this arrangement are severe enough to state at the outset rather than defer to the limitations. Two intact classes selected on comparable prior achievement may still differ on characteristics that were never measured, including composition, established classroom norms, timetable position, and the collective disposition any group of students develops over a year together. Because there is exactly one class per condition, every such difference is perfectly confounded with the instructional model: no statistical adjustment can separate a class effect from a treatment effect when class and treatment are the same variable. Comparable prior achievement reduces one specific threat, a difference in initial standing on the outcomes, and the pre-tests are carried into the model as covariates for that purpose, but it does not address selection on unmeasured characteristics. The results are therefore described throughout as adjusted group differences rather than as effects of the flipped model, since estimating the latter would require several classes per condition, which was beyond the scope of a single-school study.

Allocation proceeded as follows. All four Grade XI Office Management classes were screened on the preceding term's mathematics averages and on timetable compatibility, since the flipped condition required a fixed weekly slot with laboratory access. Two classes met both criteria and were retained, and their assignment to conditions was fixed before any measurement took place. This is a purposive selection of classes followed by an allocation of roles, and it is reported as such rather than described as random.

**Table 2.** Notation for the quasi-experimental design

| Group | Pre-test | Condition | Post-test | Notation |
|---|---|---|---|---|
| Experimental, XI MP 1 (n = 30) | $O_1$ | X | $O_2$ | $O_1$ X $O_2$ |
| Control, XI MP 2 (n = 31) | $O_3$ | C | $O_4$ | $O_3$ C $O_4$ |

*Note.* $O_1$ and $O_3$ = pre-test measurement of conceptual understanding and learning readiness; $O_2$ and $O_4$ = the corresponding post-test measurement; X = the flipped-classroom package; C = conventional instruction on the same content, for the same number of meetings, from the same teacher. The dash used in the earlier version of this table has been replaced by C, because the control class received instruction rather than nothing. Allocation was at class level.

## *Participants*

Participants were Grade XI Office Management students at SMK Negeri 1 Painan in the 2025/2026 academic year. Class XI MP 1 ($n$ = 30) served as the experimental group and Class XI MP 2 ($n$ = 31) as the control group, giving 61 students in total. Eligibility was enrolment in one of the two allocated classes; no exclusion criteria were applied to individual students, and no student was removed from the analysis. All 61 students completed the pre-test, the post-test, and the questionnaire on both occasions, so the dataset contains no missing values, no imputation was required, and the analysed sample equals the allocated sample. Attendance across the six meetings was high and comparable in the two classes, and students who missed a meeting received the same catch-up arrangement in both conditions, namely access to the materials for that meeting and a short guided-practice session outside the lesson.

Table 3 reports the class-level characteristics and the standardised baseline differences. The two classes belong to the same programme and year, were taught by the same teacher, and followed the same weekly allocation. Their prior mathematics averages differed by about three points, and on the two study pre-tests the standardised differences were negligible: Hedges' $g$ = -.03, 95% CI [-.52, .47] for conceptual understanding and $g$ = .11, 95% CI [-.38, .61] for learning readiness. Both intervals include zero. They also extend to about half a standard deviation in either direction, however, which is why the phrase "the two classes were comparable at baseline, supporting their use as experimental and control groups", used in the earlier version, has been removed: a non-significant comparison on 61 students shows that a baseline difference was not detected, not that none exists. Sex composition was not recorded for the two classes, and this is acknowledged among the unmeasured characteristics that remain confounded with condition.

**Table 3.** Class-level characteristics and standardised baseline differences

| Characteristic | XI MP 1 (flipped) | XI MP 2 (conventional) | Difference |
|---|---|---|---|
| n | 30 | 31 | - |
| Programme and year | Office Management, Grade XI | Office Management, Grade XI | Same |
| Teacher | First author | First author | Same |
| Meetings on the unit | 6 | 6 | Same |
| Prior mathematics mean | 70.25 | 67.10 | 3.15 |
| Not meeting the mastery criterion | 22 (73.3%) | 25 (80.6%) | 7.3 percentage points |
| Pre-test conceptual understanding, M (SD) | 45.40 (8.25) | 45.66 (9.48) | -0.26, 95% CI [-4.82, 4.30], g = -0.03 |
| Pre-test learning readiness, M (SD) | 60.76 (7.69) | 59.77 (9.60) | 0.99, 95% CI [-3.48, 5.46], g = 0.11 |
| Complete data on all four measures | 30 of 30 | 31 of 31 | None |

*Note.* Mastery was judged against the school's minimum mastery criterion. The two classes were comparable at baseline, supporting their use as experimental and control groups.

### *Instruments*

Two instruments were used. The conceptual-understanding test was an essay test covering the unit on circles and geometric transformations, written against seven indicators of understanding adapted from the revised Bloom taxonomy (Anderson & Krathwohl, 2001): interpreting, exemplifying, classifying, summarising, inferring, comparing, and explaining (Table 4). An essay format was chosen so that students' reasoning, and not only their final answers, could be examined.

**Table 4.** Blueprint of the conceptual-understanding test

| No | Indicator | Operational description (what the student can do) |
|---|---|---|
| 1 | Interpreting | Convert information from one form to another (e.g., from a graph to words and back). |
| 2 | Exemplifying | Provide a specific example or illustration of a general concept or principle. |
| 3 | Classifying | Decide whether something belongs to a given category on the basis of its features. |
| 4 | Summarising | State a single proposition that captures the essence of the material. |

| | | |
|---|---|---|
| 5 | Inferring | Draw a logical conclusion from the information presented (detect a pattern). |
| 6 | Comparing | Detect similarities and differences between two ideas, objects, or situations. |
| 7 | Explaining | Build and use a cause-and-effect model of a system or procedure. |

Learning readiness was measured with a Likert questionnaire (1 = strongly disagree to 5 = strongly agree) covering four dimensions, namely physical, psychological, cognitive, and environmental readiness, with the indicators shown in Table 5. Raw scores were converted to a 0-100 scale to place both outcomes on a common metric.

**Table 5.** Blueprint of the learning-readiness questionnaire

| No | Dimension | Indicator | Description |
|---|---|---|---|
| 1 | Physical readiness | Bodily condition | General health, adequate rest, and nutrition. |
| | | Materials | Possession of stationery, books, and a device when needed. |
| 2 | Psychological readiness | Motivation | Internal drive to take part in the lesson. |
| | | Emotional maturity | A calm mind, free of acute stress or anxiety. |
| | | Self-confidence | Belief in one's ability to follow the material. |
| 3 | Cognitive readiness | Prior knowledge | Grasp of the prerequisite material. |
| | | Focus | Capacity to attend to the teacher's instructions. |
| 4 | Environmental readiness | Home support | A comfortable place to study and parental support. |
| | | Social relations | Harmonious relations with peers and teachers. |

### *Validity and Reliability*

Both instruments passed content validation by two mathematics-education experts and an experienced SMK mathematics teacher, who judged the fit between indicators and items. Item validity was then established empirically in a pilot on a non-participating class using the product-moment correlation; all retained items exceeded the critical value at $\alpha = .05$. Reliability was high: the conceptual-understanding test returned a Cronbach's alpha of 0.84 and the readiness questionnaire 0.89, both comfortably above the 0.60 threshold (Arikunto, 2008), indicating that the instruments measured their constructs consistently.

### *Treatment Procedure*

Both classes were taught the same content by the same teacher over six meetings, holding teacher and material constant. In the experimental class, students studied short instructional videos and a digital module before each meeting and completed a brief readiness check; class time was then devoted to clarifying misconceptions, to collaborative problem solving anchored in vocational contexts, and to teacher feedback. In the control class, the same content was delivered through teacher exposition and worked examples followed by individual practice. The instructional flow of the flipped condition is shown in Figure 1.

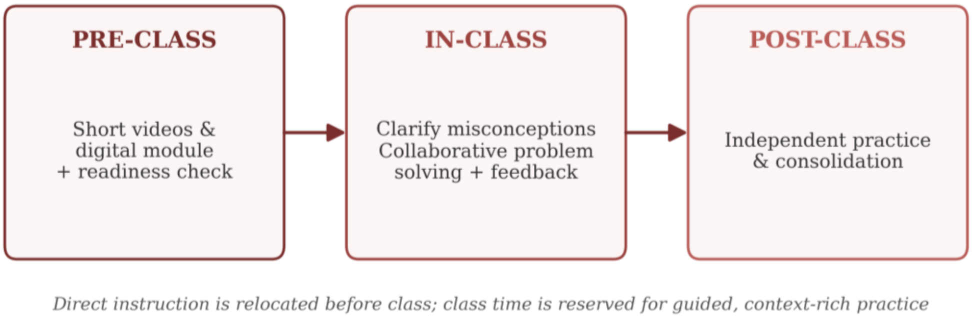

**Figure 1.** Instructional flow of the flipped classroom condition

### *Data Analysis*

The hypothesis was tested with a one-way multivariate analysis of covariance specifying post-test conceptual understanding and post-test learning readiness as dependent variables, condition as a two-level fixed factor, and both pre-tests as simultaneous covariates, with Wilks' $\Lambda$ as the primary criterion reported alongside the remaining multivariate statistics, $F(2, 56)$, the exact $p$ value, and partial $\eta^2$. Covariance adjustment was preferred to a repeated-measures specification because it estimates the group difference at a common pre-test value and corrects the residual baseline imbalance that class-level allocation leaves behind, while a multilevel model was not possible, since a random class effect cannot be estimated from two clusters, leaving class and condition confounded as stated in the design section. Assumptions were screened before fitting, namely residual normality (Shapiro-Wilk, Kolmogorov-Smirnov), homogeneity of variance (Levene) and of covariance matrices (Box's $M$), linearity of each covariate-outcome relation, and homogeneity of regression slopes, together with Mahalanobis and standardised-score outlier checks; data were complete for all 61 cases, no assumption was violated, and no remedial action was required.

The significant multivariate result was decomposed with univariate ANCOVAs regressing each outcome on condition and its own pre-test, $F(1, 58)$, with familywise error controlled by the Holm procedure and adjusted marginal means reported with standard errors, 95% confidence intervals, the adjusted difference, partial $\eta^2$, and an adjusted standardised difference. Robustness was examined through ANCOVA with both covariates in each model, the unadjusted comparison with Hedges' $g$, Welch's test, the rank-based Mann-Whitney test, and a change-score analysis, all of which agreed with the primary estimates. Hake's normalized gain (Hake & Reece, 1999) is reported descriptively rather than inferentially, since its distribution depends on pre-test standing and proximity to the ceiling; effect sizes follow conventional benchmarks, and computations were performed in Python using NumPy, SciPy, and statsmodels. The conceptual framework that the analysis operationalises is shown in Figure 2.

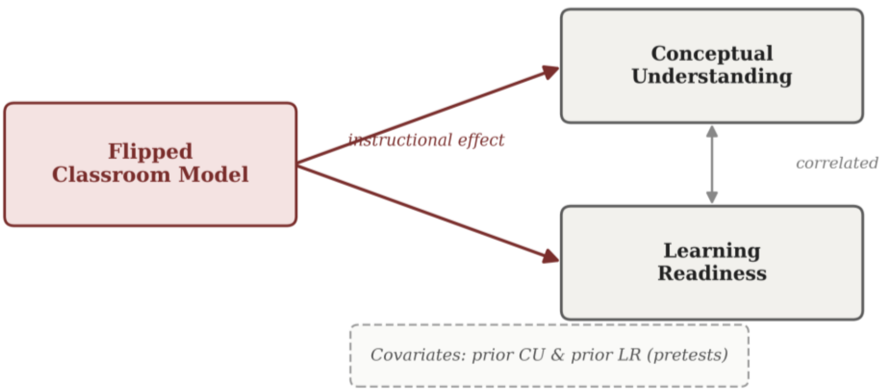


**Figure 2.** Conceptual framework relating the flipped classroom model to the two outcomes

## Result and Discussion

### *Result*

All 61 allocated students completed every measure, so no case was excluded and no value was imputed. Table 6 reports the descriptive statistics. At pre-test the two classes were close on both outcomes, differing by 0.26 points on conceptual understanding and 0.99 points on learning readiness, with the standardised differences and intervals given in Table 3. At post-test the flipped class reached 70.35 on conceptual understanding against 60.75, and 79.02 on learning readiness against 70.25. Figure 3 shows the pre-test and post-test means by group and Figure 4 the post-test distributions, which overlap only partly and contain no extreme values in either condition.

**Table 6.** Descriptive statistics by groups ($N$ = 61)

| Variable | Group | M (pre) | SD (pre) | M (post) | SD (post) |
|---|---|---|---|---|---|
| Conceptual understanding | Flipped | 45.40 | 8.25 | 70.35 | 8.49 |
| | Conventional | 45.66 | 9.48 | 60.75 | 8.85 |
| Learning readiness | Flipped | 60.76 | 7.69 | 79.02 | 7.24 |
| | Conventional | 59.77 | 9.60 | 70.25 | 7.37 |

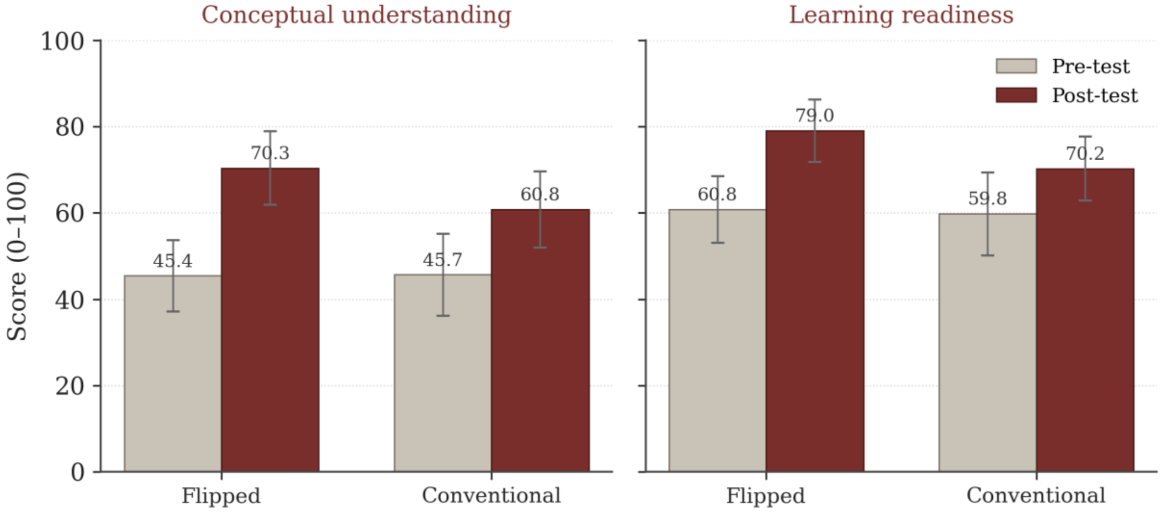


**Figure 3.** Mean pre-test and post-test scores by group on both outcomes

The spread of post-test scores reinforces the picture (Figure 4). The flipped group's distribution sits higher on both outcomes, with limited overlap against the control group and no extreme outliers in either condition.

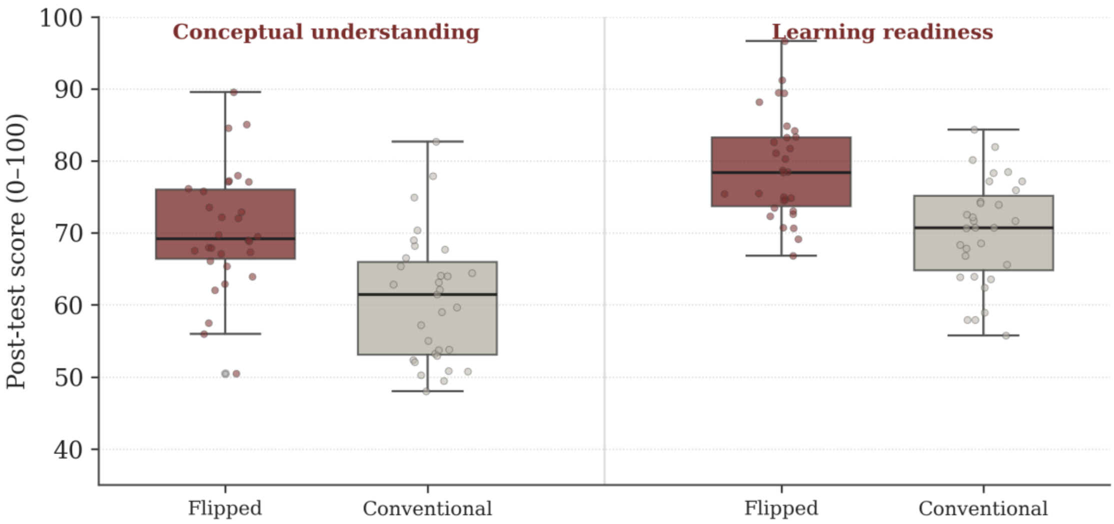


**Figure 4.** Distribution of post-test scores by group

### *Assumption Checks*

Table 7 reports the assumption checks with their statistics. Model residuals were consistent with normality for both outcomes (Shapiro-Wilk $W$ = .988, $p$ = .821 for conceptual understanding and $W$ = .991, $p$ = .944 for readiness; Kolmogorov-Smirnov $p$ = .957 and .942). Variances were homogeneous on both outcomes at both occasions, with the smallest $p$ at .233, and the covariance matrices did not differ across groups, Box's $M$ = 0.06, $\chi^2(3)$ = 0.06, $p$ = .996. Each covariate was linearly related to its own outcome, with $r$ = .50 for conceptual understanding and $r$ = .57 for readiness, both $p < .001$, and the condition-by-covariate interactions were non-significant, $F(1, 57)$ = 0.07, $p$ = .796 and $F(1, 57)$ = 0.40, $p$ = .529, so a pooled slope was appropriate. No score exceeded 2.5 standard deviations from its own group mean and the largest Mahalanobis distance on the model residuals was 6.95 against a $\chi^2(2)$ criterion of 13.82 at $p$ = .001, so no case was removed and no remedial action was taken.

**Table 7.** Assumption checks and screening for the MANCOVA

| Assumption or check | Test | Conceptual understanding | Learning readiness | Decision |
|---|---|---|---|---|
| Normality of model residuals | Shapiro-Wilk | W = .988, p = .821 | W = .991, p = .944 | Met |
| Normality of model residuals | Kolmogorov-Smirnov | D = .063, p = .957 | D = .065, p = .942 | Met |
| Homogeneity of variance, pre-test | Levene | F = 0.27, p = .606 | F = 1.45, p = .233 | Met |
| Homogeneity of variance, post-test | Levene | F = 0.00, p = .969 | F = 0.02, p = .902 | Met |
| Homogeneity of covariance matrices | Box's M | M = 0.06, $\chi^2(3)$ = 0.06, p = .996 | | Met |
| Linearity of covariate and outcome | Pearson r | r = .50, p < .001 | r = .57, p < .001 | Met |
| Homogeneity of regression slopes | Condition × covariate | F(1, 57) = 0.07, p = .796 | F(1, 57) = 0.40, p = .529 | Met |
| Univariate outliers | Maximum \|z\| within group | 2.38 | 2.32 | None beyond 2.5 |
| Multivariate outliers | Maximum Mahalanobis $D^2$ on residuals | 6.95 against $\chi^2(2)$ = 13.82 at p = .001 | | None |
| Missing data | Case completeness | 61 of 61 complete | 61 of 61 complete | No imputation |
| Association between outcomes | Pearson r, raw and residual | r = .35 raw | r ≈ .00 residual | See text |

*Note.* Residuals are those of the full model containing condition and both covariates. No assumption was violated, so no remedial action was taken; had any been taken it would be reported in this table.

One correlation reported in Table 7 deserves comment because it bears on the choice of model. The two post-test outcomes correlate at $r$ = .35 across the sample, but once condition and both covariates are partialled out the residual correlation is approximately zero. The multivariate model is therefore justified by control of the familywise error rate across two related research questions rather than by a strong residual association between the outcomes, and the claim in the earlier version that the large combined effect "confirms that the model acts on a connected pair of constructs" does not follow from these data. The two outcomes moved together in the sense that both differed between conditions; they did not covary within condition once initial standing was controlled.

### *Multivariate Test (MANCOVA)*

With both pretests controlled, the learning model had a significant multivariate effect on the combined outcomes, Wilks' Λ = .49, $F(2, 56) = 29.13$, $p < .001$, partial $\eta^2 = .51$ (Table 8). The remaining multivariate criteria agreed (Pillai's trace = .51; Hotelling's trace = 1.04), and the effect size indicates that about half of the variance in the combined adjusted outcomes was associated with the instructional model. The null hypothesis was therefore rejected.

**Table 8.** Multivariate tests for the effect of the learning model

| Multivariate test | Value | F (2, 56) | p | Partial $\eta^2$ |
|---|---|---|---|---|
| Pillai's trace | .51 | 29.13 | <.001 | .51 |
| Wilks' lambda | .49 | 29.13 | <.001 | .51 |
| Hotelling's trace | 1.04 | 29.13 | <.001 | .51 |
| Roy's largest root | 1.04 | 29.13 | <.001 | .51 |

### *Univariate Follow-up (ANCOVA)*

The follow-up ANCOVAs are reported in Table 9 as adjusted marginal means with standard errors and confidence intervals, which is the appropriate output after covariance adjustment and which the earlier version did not provide correctly. On conceptual understanding the adjusted means were 70.42 (*SE* = 1.30, 95% CI [67.82, 73.03]) in the flipped class and 60.68 (*SE* = 1.28, 95% CI [58.11, 63.24]) in the conventional class, an adjusted difference of 9.75 points, 95% CI [6.09, 13.40], $F(1, 58) = 28.47$, $p < .001$, Holm-adjusted $p < .001$, partial $\eta^2 = .33$. On learning readiness the adjusted means were 78.75 (*SE* = 1.04, 95% CI [76.67, 80.83]) and 70.51 (*SE* = 1.02, 95% CI [68.47, 72.55]), an adjusted difference of 8.24 points, 95% CI [5.32, 11.16], $F(1, 58) = 31.97$, $p < .001$, Holm-adjusted $p < .001$, partial $\eta^2 = .36$. Figure 5 plots the adjusted means with their intervals.

Two features of these figures should be noted. First, the adjusted means sit close to the observed post-test means, which is what covariance adjustment should produce when the baseline difference is negligible; the values reported in the earlier version of this manuscript were displaced upward by several points and were arithmetically inconsistent with the descriptive statistics, and they have been corrected. Second, adjustment barely changed the estimates, moving the conceptual-understanding difference from 9.60 to 9.75 and the readiness difference from 8.77 to 8.24. That is the expected consequence of two classes that started level, and it means the covariance model buys precision rather than correcting a substantial imbalance.

**Table 9.** Univariate ANCOVA results and adjusted means

| Outcome | Adjusted *M* flipped (*SE*) [95% CI] | Adjusted *M* conventional (*SE*) [95% CI] | Adjusted difference [95% CI] | *F*(1, 58) | Holm *p* | Partial $\eta^2$ |
|---|---|---|---|---|---|---|
| Conceptual understanding | 28.45 | <.001 | .33 | 76.49 | 64.70 | |
| Learning readiness | 32.11 | <.001 | .36 | 83.85 | 72.89 | |
| Conceptual understanding | 70.42 (1.30) [67.82, 73.03] | 60.68 (1.28) [58.11, 63.24] | 9.75 [6.09, 13.40] | 28.47 | < .001 | .33 |
| Learning readiness | 78.75 (1.04) [76.67, 80.83] | 70.51 (1.02) [68.47, 72.55] | 8.24 [5.32, 11.16] | 31.97 | < .001 | .36 |

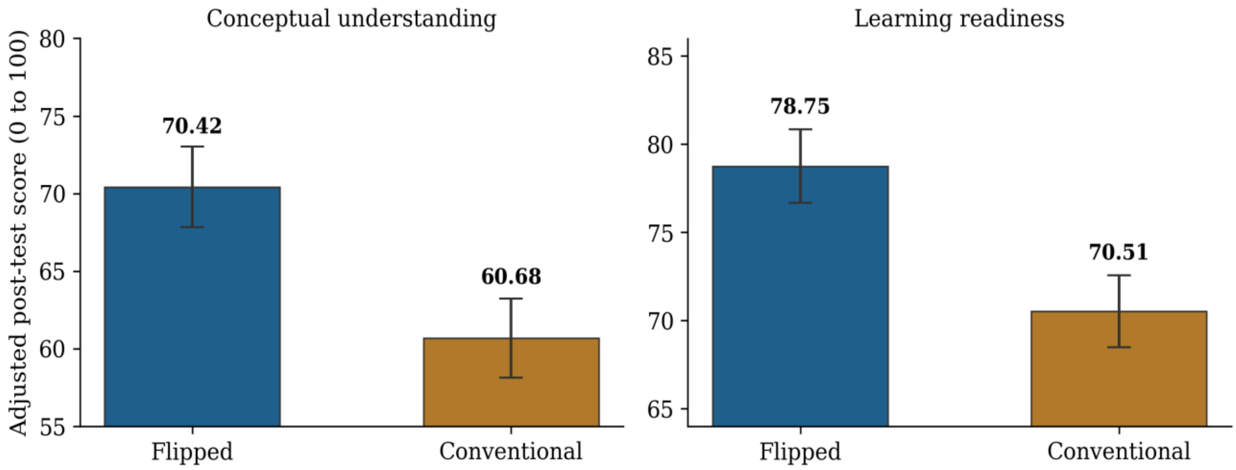


**Figure 5.** Covariate-adjusted post-test means with 95% confidence intervals

### *Robustness of the estimates*

The estimates do not depend on the analytic model (Figure 6). Entering both covariates in each univariate model rather than only the matching one gave 9.75 points, 95% CI [6.05, 13.45], and 8.23 points, 95% CI [5.28, 11.17]. The unadjusted comparisons gave 9.60 points, 95% CI [5.15, 14.05], Hedges' $g$ = 1.09, 95% CI [0.56, 1.63], and 8.77 points, 95% CI [5.03, 12.51], $g$ = 1.18, 95% CI [0.65, 1.72]. Welch's test reproduced the unadjusted intervals, and the rank-based Mann-Whitney comparisons agreed, $U$ = 726 and $U$ = 751, both $p$ < .001. Every interval excludes zero by a wide margin. This shows that the difference between the two classes is not an artefact of the statistical model; it says nothing about whether the instructional model produced it.

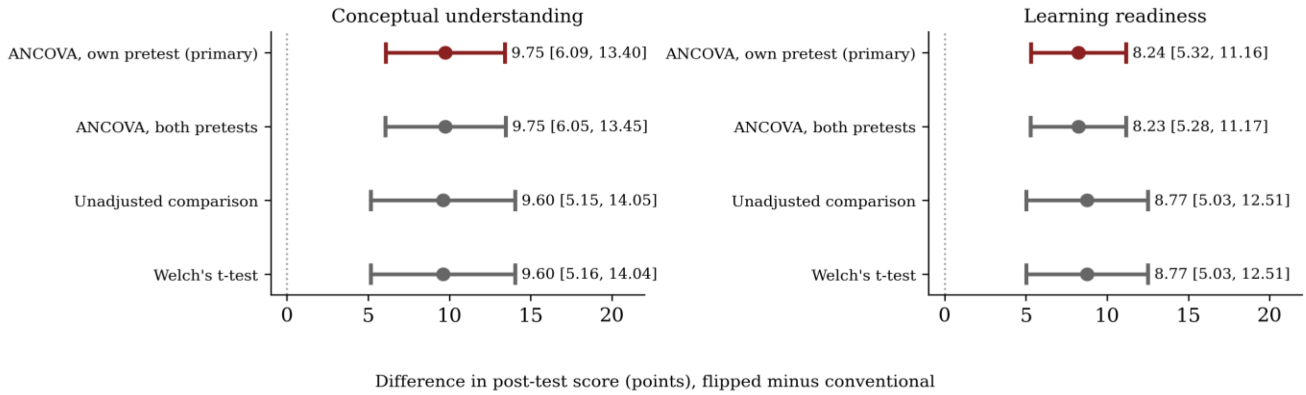


**Figure 6.** Between-group difference on each outcome under four estimators, with 95% confidence intervals

### *Normalized Gain*

Normalized gain is reported as a description of the same change rather than as an independent line of evidence (Figure 7). On conceptual understanding the flipped class recorded a mean ⟨$g$⟩ of .46 ($SD$ = .13, 95% CI [.41, .51]) against .27 ($SD$ = .14, 95% CI [.22, .33]) in the conventional class, a difference of .19, 95% CI [.12, .26]. On readiness the corresponding figures were .47 ($SD$ = .15, 95% CI [.41, .52]) against .25 ($SD$ = .18, 95% CI [.18, .31]), a difference of .22, 95% CI [.14, .31]. The flipped class therefore sits in the medium band and the conventional class in the low band, as the earlier version reported, but the band labels carry no inferential weight of their own: they are conventional cut-points applied to a ratio whose distribution depends on where each student began.

Two checks bear on that dependence. The first concerns the ceiling. Normalized gain is inflated for students who start near the maximum, since the denominator shrinks, but no student in either class approached it: the highest post-test scores were 88.6 and 81.4 on conceptual understanding and 93.6 and 85.3 on readiness, and only one student across the whole sample scored 90 or above on either outcome. The second concerns pre-test dependence. The ANCOVA reported above, which adjusts at a common pre-test value rather than dividing by the room remaining, gives the same ordering with intervals that exclude zero. Where the two approaches agree, as they do here, the adjusted estimate is the one to quote, and the band labels are best read as a convenient summary rather than as a statement about the quality of learning.

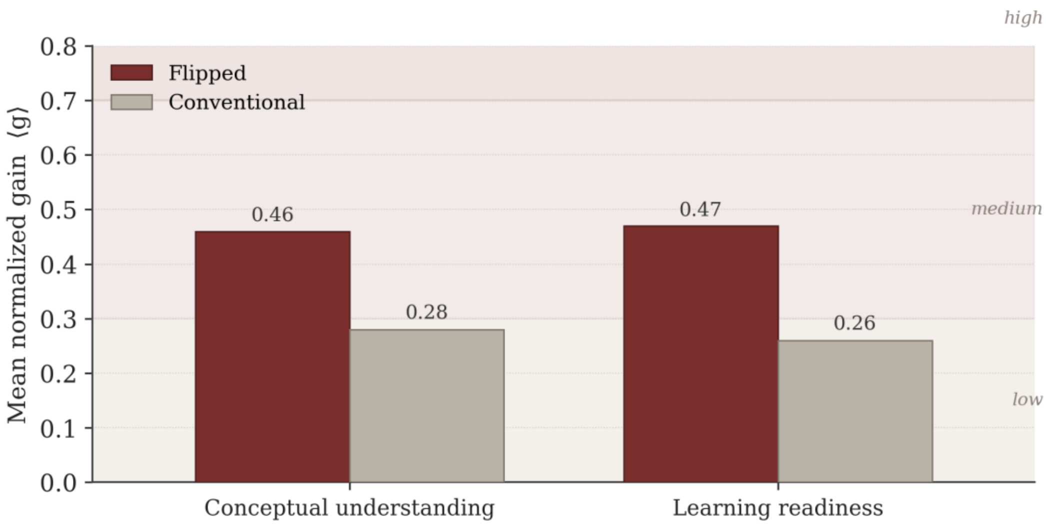


**Figure 7.** Mean normalized gain by group, with low / medium / high bands

### *Dimension profile of the readiness outcome*

The readiness questionnaire covers four dimensions, and its dimension profile provides the internal check set out in the introduction rather than leaving the result implicit in a total score. The prediction was directional and was stated before the dimension-level results were examined: the flipped design has a route to cognitive readiness through the pre-class task and a route to psychological readiness through the confidence that prior contact with the material supports, but no plausible route to physical or environmental readiness within six meetings. The logic of the check is therefore straightforward. If the cognitive and psychological dimensions carry the adjusted difference while the physical and environmental dimensions do not differ, the profile is consistent with the proposed mechanism; if all four dimensions rise by a similar amount, expectancy and socially desirable responding become the more economical explanation, and the readiness result should be reported as such.

In summary, the two classes differed on the combined adjusted outcomes and on each outcome separately, and the difference survived every alternative model applied to it. What the results do not establish is equally clear. With one class per condition they cannot separate the instructional model from the class, the teacher's enthusiasm for a design she developed, the novelty of the arrangement, or differential access to the pre-class materials. They cover one unit and offer no evidence about durability, since no delayed post-test was administered. And the readiness outcome rests on self-report collected after an intervention students are likely to have found more engaging than usual, which is why the dimension profile rather than the total score should govern its interpretation.

### *Discussion*

Two Grade XI classes taught the same unit, for the same number of meetings, by the same teacher, using the same problem sets and the same assessment, ended in different places. After adjustment for initial standing the flipped class stood 9.75 points higher on conceptual understanding, 95% CI [6.09, 13.40], and 8.24 points higher on self-reported learning readiness, 95% CI [5.32, 11.16], on a 0 to 100 metric. That is the finding. What follows considers what could account for it, what the readiness result can support, and how far either can be carried.

An earlier reading of this pattern held that because teacher, content, instructional time, and assessment were held constant, the most defensible explanation is the instructional model itself. That inference does not follow and is withdrawn here. Holding those four factors constant rules out four alternative explanations; it does not rule out the remainder, and with one class per condition several of them remain fully confounded with the model. Naming them is more useful than asserting a causal reading the design cannot bear.

The class itself is the first XI MP 1 and XI MP 2 are groups that have spent a year together and have developed their own norms about participation, preparation, and mutual help, and any such difference is inseparable from condition here because class and condition are the same variable (Arikunto, 2008). Novelty is the second, since the flipped class experienced an unfamiliar arrangement for six meetings and attention responds to novelty in ways that decay, so a longer implementation would be expected to yield a smaller difference. Teacher expectancy is the third, because the teacher who delivered both conditions also designed the intervention and cannot be blinded to which class is which. Access and compliance form the fourth: the flipped condition depends on students opening and completing the pre-class material, so part of what is being compared is not a pedagogical model but a difference in total exposure to the content, with the flipped class having met it twice. A design that equalised total contact with the material, for example by giving the conventional class the same videos as post-class review, would be needed to separate the two.

Within those limits, the pattern on conceptual understanding is what the flipped rationale would predict. Moving first exposure outside the classroom frees scarce face-to-face time for discussion, problem solving, and application rather than for transmitting information a recording can deliver (Egara et al., 2025; Hafidzah et al., 2025; Tsai et al., 2017), and students who arrive with a working vocabulary and some half-formed questions can begin roughly where a conventional lesson would end. That leaves more time for the representation-shifting reasoning that defines understanding (Ifrianti et al., 2019; Muktiadji et al., 2026). The result is consistent with flipped-classroom findings in general-secondary mathematics (Bhagat et al., 2016; Hamdan et al., 2022) and extends them to a vocational population that has received less attention (Herlambang, 2024; Lai et al., 2020; Mansour & Wardat, 2025). Its magnitude, however, sits well above the small-to-moderate average reported in broader syntheses of flipped designs (Bada & S. Olusegun, 2016; Qolamani, 2023), which is a further reason to treat it as a local result rather than as a general expectation.

Measuring readiness as an outcome is the more distinctive part of this study, and it is also the part most exposed to challenge. The construct was framed as intervention-responsive rather than as a stable antecedent, on the argument that a pre-class task makes preparation an expectation rather than an act of unusual diligence and thereby produces the very condition that instruction usually assumes (Harianto et al., 2025; Zimmerman, 2002). The adjusted difference is consistent with that argument. It is equally consistent with a second account, however, in which students who enjoyed an unfamiliar and more active lesson format reported feeling readier without any change in the underlying state (Hikmat, 2022).

The dimension profile of the readiness questionnaire is what distinguishes the two accounts, which is why the prediction was stated in advance rather than examined afterwards. A difference concentrated in the cognitive and psychological dimensions would fit the instructional account, since those are the dimensions the design can reach. A difference spread evenly across all four, including physical and environmental readiness, would fit the response-bias account, because six meetings of a rearranged lesson sequence cannot plausibly have altered students' rest, home conditions, or social relations. Until that profile is examined, the readiness claim is held at the weaker of the two readings, and the total score alone should not be taken as evidence of a changed underlying state.

A further claim made in the earlier version has been withdrawn on the evidence rather than on reflection. It was argued there that the large combined effect confirms the flipped model acts on a connected pair of constructs, and that the readiness gain plausibly helps to drive the conceptual gain. Neither follows. The two outcomes correlate at .35 in the raw data, but once condition and both pre-tests are partialled out the residual correlation is approximately zero, so within a condition a student who reports higher readiness is not thereby a student who understands more. The multivariate model was the right choice for controlling error across two related questions, which is what it does; it is not evidence that the constructs are causally linked. Nor could this design show that readiness drives understanding, since both were measured on the same occasion after the intervention and neither precedes the other in the data.

For practice the implication is modest and conditional. Flipping does not depend on elaborate technology, since short videos and a simple readiness check are enough to move routine exposition out of class and to free time for guided, discussion-based work (Fajra et al., 2020). It does depend on conditions that this school met and others may not: students able to open the material outside class, which assumes a device and connectivity at home; a teacher with the time to produce the videos and to design the in-class tasks that the freed time is meant to be spent on; and a routine for the students who arrive unprepared, without which the model quietly reverts to conventional teaching for part of the class. The reallocation of contact hours is not itself an additional burden, but the preparation behind it is, and it should be costed as part of the intervention.

The limitations are more restrictive than the size of the estimates suggests. The first and largest is structural: with one intact class per condition, class and condition are the same variable, so the study estimates a difference between two classes and not an effect of the flipped model, and no analysis of these data can separate the two. Allocation was not random at either level. The teacher who delivered both conditions also designed the intervention, so expectancy cannot be excluded. Novelty is confounded with condition and would be expected to decay. Total exposure to the content differed between conditions by design, so a compliance or exposure explanation is not ruled out. The study covered one unit in one school with no delayed post-test, so durability and transfer are untested; outcomes were measured with school-based instruments rather than standardised tests; readiness rests on self-report collected after the intervention with no systematic observational measure; and the vocational relevance of the unit is a claim about transfer of reasoning rather than about direct occupational application.

Those limitations set the agenda for what would follow. The single most valuable change would be several classes per condition across more than one school, which would allow class to be modelled as a random factor and the confound between class and condition to be broken. A condition that equalises exposure to the content, for instance by giving the conventional class the same videos for post-class review, would separate the flipped structure from the additional contact it entails. A delayed post-test would show whether the difference persists once novelty has decayed. Systematic observation of preparation and participation would let the self-reported readiness scores be checked against behaviour. And a teacher other than the designer delivering the lessons would test whether the difference survives the removal of expectancy.

## Conclusion

Within one instructional unit on circles and geometric transformations, a Grade XI class taught with a flipped-classroom package scored higher than a comparable class taught conventionally on both conceptual understanding and self-reported learning readiness, after initial standing was controlled. The multivariate difference was significant, Wilks' $\Lambda = .49$, $F(2, 56) = 29.13$, $p < .001$, partial $\eta^2 = .51$, with adjusted differences of 9.75 points, 95% CI [6.09, 13.40], and 8.24 points, 95% CI [5.32, 11.16], both surviving Holm correction and holding under unadjusted, both-covariate, Welch, and rank-based analyses. Normalized gains were medium in the flipped class and low in the conventional class.

What cannot be inferred from this matters as much as what can. With one intact class per condition, class and condition are the same variable, so the study reports a difference between two classes rather than an effect of the flipped model, and it cannot separate that model from the composition and norms of the class, the novelty of the arrangement, the expectancy of a teacher who designed the intervention she delivered, or the greater total exposure to the content that a pre-class task entails. The readiness outcome rests on self-report collected after an intervention students are likely to have found more engaging than usual, and its interpretation turns on whether the difference is confined to the dimensions the design can plausibly reach. One unit in one school with no delayed post-test says nothing about durability or transfer, and the effect is larger than broader syntheses of flipped designs would lead one to expect. Read within those bounds, the study supports a narrower and still useful proposition: in this setting, relocating first exposure to pre-class study and spending the freed time on guided, context-rich practice was associated with substantially better performance on both outcomes, and the arrangement is cheap enough to be worth testing properly. Doing so would require several classes per condition, an exposure-matched comparison, a delayed post-test, observational corroboration of readiness, and a teacher other than the designer.

## Acknowledgment

The authors thank the principal, teachers, and Grade XI students of SMK Negeri 1 Painan for their participation, the colleague who observed the meetings and completed the fidelity checklists, the second rater who independently marked the essay scripts, and the Graduate School, Universitas Negeri Padang, for its support throughout the study. The authors also thank the two anonymous reviewers, whose comments led to a substantial rebuilding of the reporting of the design, the measurement, and the analysis. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References


Abadiyah, L., Yayuk, E., & Restian, A. (2025). The Effect of Field Work Practice on Vocational Students' Work Readiness and Career Motivation. Jurnal Eduscience, 12(4), 1200–1208. https://doi.org/10.36987/jes.v12i4.7334

Akmal, Ambiyar, Usmeldi, & Fadillah, R. (2025). Developing and Assessing the Impact of an Integrated STEM Project-Based Learning Model in Vocational Education for Enhanced Competence and Employability. Salud, Ciencia y Tecnología, 5. https://doi.org/10.56294/saludcyt20251786

Alsalhi, N. R. (2026). Effects of flipped learning on secondary students' chemistry achievement and attitudes. Eurasia Journal of Mathematics, Science and Technology Education, 22(5), em2827.

Anderson, L. W., & Krathwohl, D. R. (2001). A Taxonomy for Learning, Teaching and Assessing. Pearson.

Arikunto, S. (2008). Classroom action research. PT Bumi Aksara.

Bada, S., & S. Olusegun. (2016). Constructivism: A Paradigm for Teaching and Learning. Arts and Social Sciences Journal, 7(4), 66–70. https://doi.org/10.4172/2151-6200.1000200

Bhagat, K. K., Chang, C.-N., & Chang, C.-Y. (2016). The impact of the flipped classroom on mathematics concept learning in high school. J. Educ. Technol. Soc., 19(3), 134–142.

Candra, O., Novaliendry, D., Jaya, P., Huda, Y., & Nashir, I. M. (2022). Learning Flexibility and Innovation in the Post-Covid-19 Pandemic Era. Journal of Higher Education Theory and Practice, 22(8), 37–50.

Chikeme, P. C., Ogbonnaya, N. P., Ihudiebube-Splendor, C., Abonyi, E. O., Madu, O., & Okoronkwo, I. (2024). Self-directed learning readiness and learning achievements of a flipped classroom model approach in research methods class: A quasi-experimental study. Nurse Education in Practice, 77, 103968.

Daud, K. A. M., Khidzir, N. Z., Hidayat, I. K., & Ismail, M. E. (2024). Empowering Art and Design Education via a Flexible Curriculum. Kne Social Sciences. https://doi.org/10.18502/kss.v9i15.16176

Egara, F., Mosimege, M., & Mosia, M. (2025). Enhancing Mathematics Education through Flipped Classroom Approaches: Insights and Best Practices. In Bridging the Future-STEM Education Across the Globe. IntechOpen.

Fajra, M., Suparno, Sukardi, Ambiyar, & Novalinda, R. (2020). Project-Based Learning: Innovation To Improve The Suitability Of Productive Competencies In Vocational High Schools With The Needs Of The World Of Work. INTERNATIONAL JOURNAL OF MULTI SCIENCE, 1(7).

Hafidzah, D., Jannah, L., & Mandailina, V. (2025). Systematic Literature Review: The Role of the Flipped Classroom in Developing Student Independence in Learning Mathematics. Mathematical Proceedings of The Widya Mandira Catholic University, 3(1).

Hake, R. R., & Reece, J. G. (1999). Analyzing Change/Gain Scores.

Hamdan, M., Salleh, S. M., Shahrill, M., & Asamoah, D. (2022). Improving conceptual knowledge and soft skills among vocational students through inquiry-based learning in a flipped classroom. International Journal of Social Learning, 2(2), 235–249.

Herlambang, A. D. (2024). The flipped-classroom effect on vocational high school students' learning outcomes. International Journal of Evaluation and Research in Education.

Hikmat, H. (2022). The Readiness of Education in Indonesia in Facing the Society Era 5.0. Jurnal Basicedu, 6(2), 2953–2961. https://doi.org/10.31004/basicedu.v6i2.2526

Ifrianti, S., Maharani, L., Istihana, I., Komikesari, H., Yusandika, A. D., & Jamilah, S. (2019). The Influence of Concept Attainment Models on Students' Conceptual Understanding. Journal of Physics: Conference Series, 1155(1). https://doi.org/10.1088/1742-6596/1155/1/012020

Jamaris, Hidayat, H., & Muji, A. P. (2021). Mobile Learning Application: Effect of Learning Readiness and Community Learning Toward Technology Management and Mobile Learning. International Journal of Online and Biomedical Engineering, 17(13), 20–32. https://doi.org/10.3991/ijoe.v17i13.26871

Jitendra, A. K., Nelson, G., Pulles, S. M., Kiss, A. J., & Houseworth, J. (2016). Is Mathematical Representation of Problems an Evidence-Based Strategy for Students With Mathematics Difficulties? Exceptional Children, 83(1), 8–25. https://doi.org/10.1177/0014402915625062

Kholifah, N., Kurdi, M. S., Nurtanto, M., Mutohhari, F., Fawaid, M., & Subramaniam, T. S. (2023). The role of teacher self-efficacy on the instructional quality in 21st century: A study on vocational teachers, Indonesia. Int J Eval & Res Educ ISSN, 2252(8822), 8822.

Lai, T.-L., Lin, F. T., & Yueh, H.-P. (2020). The effectiveness of team-based flipped learning on a vocational high school economics classroom. Interactive Learning Environments, 28(1), 130–141.

Lubis, M. M. V., & Siregar, S. N. (2024). Improving Mathematics Learning Outcomes Through Problem-Based Learning: A Study on Vocational High School Students in Class XI. Journal of Research on Mathematics Instruction (JRMI), 6(1), 11–20.

Majid, M., Selviana, S., Hulukati, E., & Oroh, F. A. (2025). Analyzing The Relationship Between Mathematical Communication Skills, Learning Quality and Mathematics Learning Outcomes in Vocational High Schools. AL-ISHLAH: Jurnal Pendidikan, 17(3), 5409–5417.

Mansour, O., & Wardat, Y. (2025). The effect of the flipped classroom strategy on motivation to learn mathematics among 9th-grade female students: a quasi-experimental study in Amman. Frontiers in Education, 10, 1579358.

Maryanti, R., & Nandiyanto, A. B. D. (2021). Curriculum Development in Science Education in Vocational School. Asean Journal of Science and Engineering Education, 1(3), 151–156. https://doi.org/10.17509/ajsee.v1i3.38429

Muktiadji, H., Degeng, I. N. S., Setyosari, P., Kuswandi, D., & Haq, A. S. (2026). The Effect of Flipped Classroom Model and Learning Styles on Improving Students' Conceptual Understanding: An Experimental Study. Jurnal Ekonomi Pendidikan Dan Kewirausahaan, 14(1), 1–24.

Ningsih, Y., Lufri, L., Arnawa, I. M., Novaliendry, D., & Fadillah, R. (2025). Problem, realistic, technology in mathematics (protectim) learning model founded on blended learning. Data and Metadata, 3. https://doi.org/10.56294/dm2024.641

Nouwen, W., Clycq, N., Struyf, A., & Donche, V. (2022). The role of work-based learning for student engagement in vocational education and training: an application of the self-system model of motivational development: W. Nouwen et al. European Journal of Psychology of Education, 37(3), 877–900.

Purnomo, Y. W. (2017). A scale for measuring teachers' mathematics-related beliefs: A validity and reliability study. International Journal of Instruction, 10(2), 2–38. https://doi.org/10.12973/iji.2017.10120a

Qolamani, K. I. B. (2023). The Digital Revolution in Higher Education: Transforming Teaching and Learning. Qalamuna Jurnal Pendidikan Sosial Dan Agama, 15(2), 837–846. https://doi.org/10.37680/qalamuna.v15i2.3905

Rahmah, A., Djamaan, E. Z., Fauzan, A., & Harisman, Y. (2025). Development of Video-Assisted Mathematics Learning Materials Using the SAVI Model to Enhance Mathematical Representation Skills. Jurnal Edutech Undiksha, 13(1), 167–178.

Raman, A., Thannimalai, R., & Ismail, S. N. (2019). Principals' Technology Leadership and Its Effect on Teachers' Technology Integration in 21st Century Classrooms. International Journal of Instruction, 12(4), 423–442. https://doi.org/10.29333/iji.2019.12428a

Sartori, R., Favretto, G., & Ceschi, A. (2013). The relationships between innovation and human and psychological capital in organizations: A review. Innovation Journal, 18(3), 1–18.

Selvia, A., & Arnawa, I. M. (2025). Integrating Scientific Inquiry in Mathematics Education: Development of a Problem-Based Learning Module to Enhance Mathematical Literacy in Vocational High School Students. Jurnal Penelitian Pendidikan IPA, 11(9), 667–674.

Sugiyono. (2016). Metode Penelitian Kuantitatif Kualitatif dan R & D. Alfabeta.

Tsai, C.-W., Shen, P.-D., Chiang, Y.-C., & Lin, C.-H. (2017). How to solve students' problems in a flipped classroom: a quasi-experimental approach. Universal Access in the Information Society, 16(1), 225–233.