Jurnal Pustaka Indonesia (JPI)
2026, Vol. 6, No. 2, 305–323
https://doi.org/10.62159/jpi.v6i2.2543

OPEN ACCESS Research Article

Seeing the solution: video-assisted direct instruction, learning engagement, and mathematics achievement in vocational education

Yulia Cresna[1*], Abna Hidayati[2], Fetri Yeni J.[3], and Hani Fannisa[4]
[1-4] Universitas Negeri Padang, Padang, Indonesia

Abstract
Low participation has long shadowed mathematics instruction in Indonesian vocational high schools, where teacher-led delivery leaves learners watching rather than thinking. This study tested whether delivering the demonstration phase of the Direct Instruction (DI) model through purpose-built videos raises learning engagement and mathematics achievement, and explored how the two outcomes are related. A quasi-experiment with a non-equivalent pre-test-post-test control-group design was conducted with 58 tenth-grade students at a vocational school in West Sumatra, allocated to video-assisted DI (n = 30) or conventional DI (n = 28). Videos followed a documented design specification, and implementation fidelity, viewing compliance, and instructional time were recorded in both classes. Because a non-significant baseline test is not evidence of equivalence, the primary analysis was an analysis of covariance on post-test achievement with the pre-test as covariate, with familywise error controlled across both confirmatory outcomes. Engagement was higher in the experimental class by 10.92 points, 95% CI [7.90, 13.94], $t(56) = 7.25$, Hedges' $g = 1.88$. Adjusted post-test achievement was higher by 13.16 points, 95% CI [9.66, 16.67], $F(1, 55) = 56.73$, partial $\eta^2 = .51$, and remained stable across separate-slopes, Welch, and rank-based analyses. Engagement and achievement were strongly associated, $r = .74$, 95% CI [.60, .84], though this fell to .43 and .62 within classes. An exploratory decomposition attributed roughly half the adjusted group difference to engagement; because engagement was measured once, concurrently with the post-test, this is a statistical partition rather than a causal mechanism. Findings from a single school require replication.

ARTICLE HISTORY
Received : 13 April 2026
Revised : 21 June 2026
Accepted : 24 July 2026

KEYWORDS
Direct instruction; learning video; student engagement; mathematics achievement; vocational education

CORRESPONDING AUTHOR
***Yulia Cresna**, Universitas Negeri Padang, Padang, Indonesia. Email: yuliayc86@gmail.com

Introduction

Mathematics achievement in Indonesian vocational high schools sits at the intersection of two well-documented problems: a curriculum that requires mathematical reasoning as the foundation of technical competence, and a delivery routine that leaves most learners in the role of spectators (Suharno et al., 2020; Widana et al., 2021). This study addresses the second problem while measuring both. It asks whether replacing the teacher-talk portion of a structured instructional model with purpose-built video changes how much students take part and how much they learn, and it examines how those two things are related. The remainder of this introduction sets out the theoretical account that motivates the design, reviews what the evidence does and does not establish, and states the questions and their limits.

The problem is concrete at the study site. Odd-semester examination records for the 2025/2026 cohort show that of 243 tenth-grade students, only 83 (34.2%) reached the minimum mastery criterion of 70 while 160 (65.8%) fell short, with class averages clustered between 50 and

55. A classroom observation in one tenth-grade class corroborated the pattern. Across five engagement indicators, namely attending to explanations, joining discussions, asking questions, responding to the teacher, and working through in-class exercises, the majority of students were rated as passive, with question-asking and discussion the weakest of all. A follow-up class quiz returned the predictable result, with 21 of 30 students below the mastery threshold. Low participation and weak achievement appear together in these data (Feng & Hou, 2023; Rahmiati et al., 2024; Wannapiroon et al., 2021), although co-occurrence in a single cross-section says nothing about which produces which.

Three bodies of theory bear on the design, and it is worth stating how they connect before any empirical claim is made. The first concerns the instructional model. Direct Instruction organises a lesson into stating objectives, presenting and demonstrating content, guided practice, feedback, and independent assessment, a sequence whose effectiveness rests on managing what a novice has to hold in mind at each step (Mubai et al., 2023). Its evidence base is strong, particularly for learners who struggle with mathematics (Doabler et al., 2021; Irfan et al., 2023). Its documented weakness lies in the presentation phase: when the demonstration is delivered as continuous teacher exposition, the pacing is fixed by the teacher, the representation is transient, and the learner has no means of re-entering a step that was missed (Lee et al., 2023; Rohanai et al., 2019; Statti, 2020).

The second concerns why video might alter that specific phase. The cognitive theory of multimedia learning holds that learning improves when narration and relevant graphics are presented together, when extraneous material is removed, and when a complex procedure is segmented into learner-controlled units (McLeod, 2018). Cognitive load theory supplies the reason: working memory is the binding constraint during skill acquisition, and instructional formats that reduce extraneous load leave more capacity for the schema construction that learning consists of (Chan, 2018). Video affects the presentation phase of DI on exactly these terms. It makes a transient demonstration re-playable, so a step missed need not be a step lost; it allows segmentation, so a multi-step worked example can be paused between steps; and it fixes the visual representation, so the teacher's talk during the lesson can be spent on interpretation rather than on re-drawing. None of these is a motivational claim. They are claims about the load imposed by the format (Irfan et al., 2025; Khan et al., 2022).

The third concerns engagement, and here the theoretical argument has to be stated carefully because it is the argument the design tests least well. Engagement is standardly treated as having three components: behavioural participation, emotional involvement, and the cognitive effort a learner invests in making sense of content (Struyf et al., 2019; Zhu & Kaiser, 2022). The three components should not respond to video in the same way or for the same reasons. Emotional engagement would be expected to respond to novelty and to the perceived relevance of applied, trade-adjacent examples, and it is the component most likely to decay once the medium becomes routine (Demszky et al., 2023; Reddy et al., 2021; Romero & García, 2023). Behavioural engagement should respond to the change in what the lesson asks students to do, since time freed from copying a demonstration is time available for guided practice and discussion. Cognitive engagement is the component the multimedia account predicts most directly, because segmentation and replay give a learner the opportunity to re-process a step rather than to abandon it. If the medium were doing mainly affective work, the emotional component should move most; if it were doing pedagogical work, the cognitive component should (Akmal et al., 2025; Lutz et al., 2023; Powell et al., 2021).

The step from that account to a mechanism claim is the one this study cannot complete, and saying so at the outset is more useful than discovering it in the limitations. A mechanism claim requires that the proposed mediator be manipulated or at least measured before the outcome, and that the mediator-outcome relation be free of common causes (Bullock-Yowell & Reardon, 2024). Neither condition holds here. Engagement was measured once, after instruction, on the same occasion as the achievement post-test, so a student who did well may report having been engaged for that reason as easily as the reverse. Both measures also come from the same students at the same sitting, which invites common-method variance (Satnarine, 2023; Shelle et al., 2018). The relation between engagement and achievement is therefore treated throughout as an association observed under an experimentally manipulated instructional condition, and the decomposition reported later is a statistical partition of an adjusted group difference rather than an estimate of how much of the effect engagement carries.

The literature on video-supported instruction is large and, read selectively, uniformly encouraging. Read as a whole it is more conditional. The most comprehensive synthesis available reports that video generally improves learning relative to existing teaching, with the largest benefits where video is added to rather than substituted for existing instruction, and smaller and more variable benefits where it replaces a teacher (Muskhir et al., 2024; Noetel et al., 2021). Design features moderate the result: shorter segments hold attention better than long ones, and production choices that seem consequential often are not, with the presence of a presenter's face producing mixed and sometimes null effects on retention (Guo et al., 2023; Guzey et al., 2016). A durable methodological warning applies to this whole literature. Comparisons between a media-delivered lesson and a conventionally delivered one confound the medium with the instructional method, the content, and the novelty of the arrangement, so an advantage for the media condition does not by itself license a claim about the medium (Imaran et al., 2019; Zhou et al., 2024). Table 1 summarises the studies closest to the present design against these criteria.

Table 1. Mapping of the closest prior work against the present design

Study	Setting	How video was used	Video design features reported	Outcomes measured	What remains unaddressed
Pardimin et al. (2019)	Secondary mathematics, Indonesia	ICT media developed to support the DI method	Not reported	Instructional effectiveness (1)	No participation outcome; video design not specified
Nouwen et al. (2022)	Vocational education and training	No video; work-based learning	Not applicable	Engagement (1)	No achievement outcome; no media manipulation
Fyfield et al. (2022)	Review of instructional video design	Review of design choices	Segmentation, signalling, narration, length	Mixed across studies	Design principles catalogued but not tested inside a named instructional model
Lutz et al. (2023)	Pre-vocational programmes, Germany	Audiovisual media within structured instruction	Not reported	Mathematical skills (1)	Engagement not measured; conditions for null results not examined

Surur et al. (2023)	School mathematics, Indonesia	Video-assisted online interactive learning	Not reported	Learning outcomes (1)	Single outcome; no baseline adjustment reported
Mubai et al. (2023)	Vocational practicum	Flipped Direct Instruction	Not reported	Learning model effectiveness (1)	Medium confounded with flipped sequencing
Santiago & Nabayra (2024)	General mathematics	Video-assisted learning module	Not reported	Least-learned competencies (1)	No engagement outcome; module and video not separated
Rahmah et al. (2025)	Mathematics, SAVI model	Video-assisted learning materials	Not reported	Mathematical representation (1)	Development study; no between-group estimator
Labak et al. (2025)	Teacher professional development	Online video-based professional development	Not reported	Teaching practice and student learning (2)	Video directed at teachers, not learners
Noetel et al. (2021)	Systematic review, higher education	Video added to or substituting for teaching	Length, interactivity, addition versus substitution	Learning (meta-analytic)	Higher education; vocational school not represented
Sezgin & Yüzer (2022); Shelle et al. (2018)	Online courses	Video as the primary medium	Length, production style, presenter's face	Engagement proxies and retention (2)	Null and mixed findings on production features; no classroom instructional model
Present study	**Grade 10 vocational mathematics, Indonesia**	**Video carries the presentation and demonstration phase of DI**	**Specified in advance and audited: see Table 4**	**Engagement and achievement (2)**	**Mediator measured only after instruction; ordering not established**

Note. Outcome counts in parentheses give the number of learner outcomes measured. "Not reported" means the source does not describe the design of the video material itself, which is the modal case in this literature and a limit on its cumulative value.

Two things follow from the table. First, the studies that report positive effects of video within structured instruction rarely document the design of the video itself, so a reader cannot tell whether a null result elsewhere reflects a different medium or a different quality of material, and cannot reproduce what was done. Second, none of the studies measures a participation outcome and an achievement outcome on the same students under the same manipulation, which is what would be required even to describe how the two move together (Christensen et al., 2015; Chukwuedo et al., 2022; Kholifah et al., 2023). The present study addresses both gaps directly: it documents the video specification and the evidence that the materials met it, and it measures engagement and achievement in the same design. It does not close the third and larger gap, which is the absence of temporally ordered mediator measurement, and the analysis is framed accordingly.

This study evaluates whether Direct Instruction in which the presentation and demonstration phase is carried by purpose-built learning video produces higher learning engagement and higher mathematics achievement than Direct Instruction in which the same phase is delivered by teacher exposition, and it describes how the two outcomes are related. Two questions were designated in advance as confirmatory. (1) Does video-assisted DI raise learning engagement relative to conventional DI? (2) After adjustment for pre-test standing, does video-assisted DI raise mathematics achievement relative to conventional DI? A third question is stated as exploratory rather than confirmatory. (3) How are engagement and achievement related within this design, and how much of the adjusted group difference in achievement is statistically attributable to the difference in reported engagement?

The wording of the third question is deliberate and marks a departure from the way this issue is usually framed. An earlier version of this work asked whether engagement is the pathway through which the instructional change reaches achievement. That formulation presupposes what the design cannot deliver, namely a mediator measured before the outcome and free of common causes with it. What the present design can deliver is a description: an association between two variables, one of which was raised by a manipulated instructional condition, together with a decomposition of the group difference that shows how much of it coincides statistically with the engagement difference. That is a weaker statement than a mechanism, and it is the statement the data support. The conditions under which the stronger claim could be tested are set out in the discussion.

Method

Research design

The study used a quantitative, quasi-experimental approach with a non-equivalent pre-test-post-test control-group design, the standard choice when intact classes cannot be randomly reassembled (Creswell & Creswell, 2017; Sugiyono, 2016). Both groups sat a pre-test and a post-test on mathematics achievement. The experimental group received Direct Instruction in which the presentation and demonstration phase was carried by learning video; the control group received Direct Instruction in which the same phase was delivered by teacher exposition. In standard notation the design is O_1 X_1 O_2 for the experimental class and O_1 X_2 O_2 for the control class, where the letter between the two observations denotes the condition that class received. Table 2 sets out the layout.

The role of the pre-test in this design needs to be stated precisely, because it is easy to overstate. Allocation was at class level, so the two groups are non-equivalent by construction, and a pre-test does not repair that. In particular, a non-significant independent-samples test on pre-test scores is evidence that a baseline difference was not detected in a sample of 58, not evidence that the groups are equivalent; with this sample size the test would fail to detect standardised differences well inside the range that matters. The pre-test is therefore used in two ways: its standardised group difference is reported with a confidence interval so that the size of any residual imbalance is visible, and it is carried into the outcome model as a covariate rather than treated as a hurdle that has been cleared.

Table 2. Notation for the non-equivalent control-group design

Group	Pre-test	Condition	Post-test	Notation
Experimental (X RPL 1, n = 30)	O_1	X_1	O_2	O_1 X_1 O_2
Control (X RPL 2, n = 28)	O_1	X_2	O_2	O_1 X_2 O_2

Note. O_1 = pre-test achievement; O_2 = post-test achievement; X_1 = Direct Instruction with the presentation and demonstration phase carried by learning video; X_2 = Direct Instruction with the same phase delivered by teacher exposition. The engagement questionnaire was administered once, on the same occasion as O_2. Allocation was at class level, not at student level.

Participants and setting

The population comprised all tenth-grade students at the study school in the 2025/2026 academic year, that is, eight classes totalling 243 students. Selection proceeded in two steps. All eight classes were first screened on their odd-semester mathematics averages, and the two classes whose averages were closest to each other and whose timetables allowed the same weekly slot were retained. The two retained classes, X RPL 1 and X RPL 2, belong to the same vocational programme, Software Engineering, and were taught by the same teacher. Allocation of the two classes to conditions was then decided by a coin toss conducted before the pre-test, which assigned X RPL 1 (n = 30) to the video-assisted condition and X RPL 2 (n = 28) to the conventional condition. The unequal group sizes reflect the enrolment of the two intact classes and not any exclusion made by the researcher.

Attendance and attrition were recorded meeting by meeting. All 58 enrolled students sat the pre-test, the post-test, and the engagement questionnaire, so no case was lost and no value was imputed; the analysed sample equals the allocated sample. Attendance across the instructional meetings was high and comparable in the two classes, and no student missed more than one meeting. Students who missed a meeting received the same catch-up arrangement in both classes, namely access to the relevant materials and a guided-practice session outside the lesson.

Safeguards against expectancy and performance bias

Both classes were taught by the first author, who is also their regular mathematics teacher. The arrangement holds teacher quality constant across conditions, which is a real advantage, but it also creates an evident risk that the person who designed the treatment delivers it with more energy than the comparison. Five safeguards were applied. First, the two lessons were matched on topic, learning objectives, number of meetings, and instructional minutes, and both used the same five-phase DI sequence, so only the delivery of the presentation phase differed. Second, both lesson plans were written before implementation and reviewed by two lecturers in educational technology, and both were followed as scripts rather than improvised.

Third, an observing colleague completed a phase-by-phase fidelity checklist at every meeting in both classes, recording whether each planned action occurred and how long each phase lasted, so that any drift in time on task between conditions would be visible rather than assumed away. Fourth, students were not told which class had been designated experimental, and both were told they were taking part in a study of mathematics learning media. Fifth, the engagement questionnaire was administered under code numbers, collected by the observing colleague rather than by the teacher, and preceded by an explicit statement that responses would not affect marks; the achievement tests were scored from anonymised answer sheets. Ethical clearance was obtained before data collection began, the school principal granted permission, and students and guardians gave informed consent. These measures reduce the risk of expectancy effects; they do not remove it, since a teacher cannot be blinded to a condition he or she is delivering.

Instruments

Engagement was measured with a questionnaire built on the three established dimensions of learning engagement, namely emotional, cognitive, and behavioural (Entwistle et al., 2025; Krishnawati et al., 2023; Reddy et al., 2021), comprising 15 items, five per dimension, rated on a four-point Likert scale from 4 (strongly agree) to 1 (strongly disagree). Total scores range from 15 to 60. Content validity was judged by subject experts, and internal consistency was examined with Cronbach's α on the try-out sample, which confirmed that the total scale and each of its three dimensions met the conventional threshold for acceptable reliability. Achievement was measured with a multiple-choice mathematics test aligned to the tenth-grade curriculum, administered in parallel forms before and after instruction so that the post-test was not the pre-test re-administered. Content validity was judged by the same experts, and the try-out confirmed acceptable reliability together with item difficulty and discrimination indices falling within the ranges conventionally regarded as satisfactory.

Table 3. Instruments, dimensions, item counts, and score ranges

Instrument	Dimension	Items (k)	Response format	Score range
Engagement questionnaire	Emotional	5	Likert 1 to 4	5 to 20
	Cognitive	5	Likert 1 to 4	5 to 20
	Behavioural	5	Likert 1 to 4	5 to 20
	Total	**15**		**15 to 60**
Mathematics achievement test	Curriculum-aligned content	20	Multiple choice, parallel forms at pre-test and post-test	0 to 100

Note. Item counts, reliability coefficients, and the difficulty and discrimination ranges of the achievement test must be entered from the try-out output in the thesis before submission.

Two measurement decisions constrain what the engagement scores can support. The first is timing: the questionnaire was administered once, after instruction and on the same occasion as the achievement post-test, leaving no pre-treatment engagement score, no change score, and no temporal separation between the putative mediator and the outcome. A student who has just performed well may report having been engaged partly for that reason, so the direction of the association is not identified by the data. The second is method: engagement is self-reported while achievement is a scored test, which avoids drawing both variables from a single instrument, but both were collected from the same students at the same sitting, so shared method and mood variance remain plausible sources of the observed correlation (Rohanai et al., 2019). Anonymity, collection by someone other than the teacher, and the explicit no-grading assurance limited socially desirable responding, though they cannot eliminate demand characteristics in a class aware it has been taught with an unusual medium. No observational or platform trace measure of participation was collected, so nothing corroborates the self-reported scores from an independent source.

Procedure

Because a recommendation to use "well-designed" video is empty unless the design is specified, the materials were built to a written specification derived from the cognitive theory of multimedia learning and the review evidence on instructional video design (Fyfield et al., 2022; Mayes & Rittschof, 2021). Six videos were produced, one per meeting, covering the same worked examples and steps as the teacher exposition used in the control class. Each was divided into two to three learner-paced segments separated by a pause point and a teacher prompt, so that no unbroken explanation exceeded about four minutes; extraneous music, animation, and illustration were excluded; each step was signalled by colour-coded highlighting with the preceding step left visible; and narration accompanied a developing worked example, with on-screen text limited to symbols, numerals, and step labels. Every video presented one fully worked example before a faded second example completed in guided practice, ran six to nine minutes to match the control-class exposition time, and used a voice-over recorded by the teacher without a talking-head window, since the evidence on that feature is mixed rather than favourable. All six were audited against the specification before use by two lecturers in educational technology and one mathematics teacher.

Both classes followed the same five-phase DI sequence across six meetings of 2 × 45 minutes with the same topics, objectives, and practice problems, and the intended difference was confined to the presentation phase, carried by video in the computer laboratory in the experimental class and by teacher exposition at the board using the same script in the control class. An observing colleague logged the minutes spent in each phase, fidelity to the lesson plan, and whether each student was present for the full running time, since a video-based treatment that is not watched is not a treatment. The medium was nevertheless not the only difference between the two lessons. Video fixes the pacing of the demonstration and makes it re-playable, which is the intended change, but it also frees the teacher to circulate while it runs, moves the lesson to the laboratory, and introduces a novel arrangement whose effect on attention is by definition temporary. Time on task and the wording of the explanation were held constant or recorded, but setting and novelty were not controllable within this design, so the comparison is between two packages that differ mainly but not solely in the medium of the demonstration phase.

Data analysis

The analytic strategy was fixed before the data were examined. Two outcomes were designated confirmatory, learning engagement and mathematics achievement, with one test specified for each and familywise error controlled by the Holm procedure at $\alpha = .05$; every other analysis, including the engagement dimensions, the normalized gain, and the decomposition involving engagement, is secondary or exploratory. For achievement the confirmatory test was an analysis of covariance on post-test scores with the pre-test as covariate and condition as fixed factor, pre-specified over a post-test-only comparison because it uses rather than discards the pre-test information and adjusts for the residual baseline imbalance that class-level allocation cannot rule out and a non-significant baseline test does not exclude. For engagement no covariate was available, since the questionnaire was administered once, so the confirmatory test was an independent-samples comparison with a confidence interval and Hedges' g; this asymmetry between the two outcomes is a limitation rather than a design choice. Assumptions were checked beforehand, namely normality of all distributions and of the covariance-model residuals (Shapiro-Wilk), equality of variances (Levene's test), linearity of the covariate-outcome relation, and homogeneity of regression slopes. Data were complete for all 58 cases, outliers were screened by standardising every score within its own group and occasion, and all estimates are reported with 95% confidence intervals and effect sizes.

Because a very large point estimate from 58 students warrants more than a single model, robustness was examined in four ways: re-estimating the adjusted difference with separate slopes in the two classes, comparing it with the unadjusted independent-samples estimate, with Welch's test relaxing equal variances, and with the rank-based Mann-Whitney U test. Hake's normalized gain was computed as a descriptive summary of improvement relative to the room available for it, not as inferential corroboration, since its distributional properties depend on pre-test standing and proximity to the ceiling. For the exploratory analysis of the engagement-achievement relation, post-test achievement was regressed on condition, engagement, and the pre-test, and the product of the condition-to-engagement and engagement-to-achievement paths was taken as a statistical decomposition of the adjusted condition difference, with its confidence interval obtained by percentile bootstrap with 10,000 resamples rather than the Sobel test, which assumes normality for a generally skewed product term (Dewi et al., 2025). The engagement-achievement correlation is reported both across the whole sample and within each class, since a pooled correlation across two groups differing on both variables is inflated by the group separation itself. Computations were performed in Python using NumPy, SciPy, and statsmodels.

Result and Discussion

Result

Results are reported in five parts: data completeness and descriptive statistics, assumption checks and baseline comparability, the two confirmatory tests, the robustness checks, and the exploratory analysis of the engagement-achievement relation.

Descriptive statistics and baseline equivalence

All 58 allocated students completed the pre-test, the post-test, and the engagement questionnaire, so the dataset contains no missing values and no case was excluded or imputed. Screening for extreme values found no observation more than 2.5 standard deviations from its own group and occasion mean, so no case was trimmed or winsorised. Table 4 reports the descriptive picture. At pre-test the two classes were close together, with the control class marginally ahead. After instruction the experimental class reached a post-test mean of 76.37, above the school's mastery threshold of 70, against 63.49 in the control class, and reported engagement of 48.60 of a possible 60 against 37.68. Figure 1 shows the achievement means at both occasions.

Table 4. Descriptive statistics by group

Variable	Group	*M*	*SD*	Min	Max
Pre-test	Experimental	53.13	7.13	41.0	69.5
	Control	54.27	8.03	40.1	72.5
Post-test	Experimental	76.37	7.69	59.3	91.8
	Control	63.49	5.79	52.2	78.5
Engagement	Experimental	48.60	5.49	37	58
	Control	37.68	5.98	23	52
N-gain ⟨g⟩	Experimental	0.49	0.18	0.01	0.82
	Control	0.18	0.18	-0.43	0.42

Note. Exp = experimental group (n = 30); Control (n = 28).

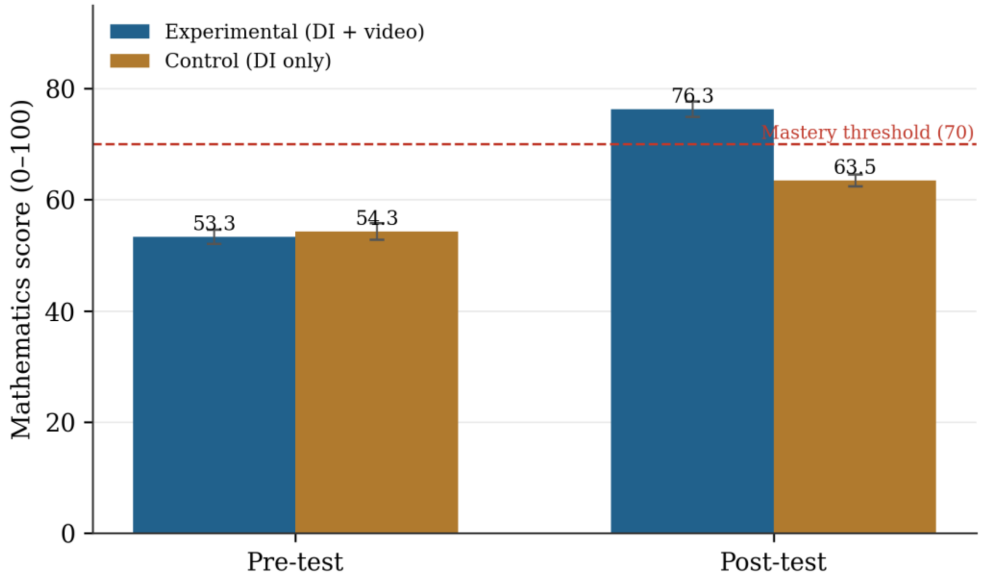

Figure 1. Mean pre- and post-test mathematics scores by group, with standard-error bars and the mastery threshold (70).

Assumption checks and baseline comparability

Assumption checks are reported in Table 5. Shapiro-Wilk tests were consistent with normality for post-test scores, engagement scores, and experimental-group normalized gains in both classes; the control-group normalized gain showed a mild departure, which is one reason the gain is treated descriptively. The residuals of the covariance model were also consistent with normality. Levene's tests found homogeneous variances for every comparison. The condition-by-covariate interaction was not significant, so the assumption of homogeneous regression slopes held and a pooled slope was appropriate.

Table 5. Assumption checks for the confirmatory analyses

Check	Statistic, Experimental	Statistic, Control	Decision
Shapiro-Wilk, pre-test	W = 0.976, p = 0.701	W = 0.972, p = 0.640	Normal
Shapiro-Wilk, post-test	W = 0.978, p = 0.778	W = 0.961, p = 0.377	Normal
Shapiro-Wilk, engagement	W = 0.962, p = 0.339	W = 0.961, p = 0.367	Normal
Shapiro-Wilk, normalized gain	W = 0.977, p = 0.747	W = 0.908, p = 0.018	Mild departure in control class
Shapiro-Wilk, ANCOVA residuals	W = 0.976, p = 0.303 (both groups)		Normal
Levene, pre-test	F = 0.30, p = 0.588		Homogeneous
Levene, post-test	F = 1.52, p = 0.223		Homogeneous
Levene, engagement	F = 0.21, p = 0.652		Homogeneous
Levene, normalized gain	F = 0.13, p = 0.720		Homogeneous
Condition × covariate interaction	F(1, 54) = 0.33, p = 0.567		Slopes homogeneous

Note. The mild departure from normality in the control-group normalized gain is one reason that measure is reported descriptively; the rank-based Mann-Whitney comparison, which makes no distributional assumption, agreed with every parametric test.

Baseline comparability was examined directly rather than inferred from a significance test. The experimental class scored 1.14 points lower at pre-test, 95% CI [-5.13, 2.85], $t(56) = -0.57$, $p = 0.569$, Hedges' $g = -0.15$. The interval includes zero and the standardised difference is small, but the interval also extends to about five points in either direction, which is why the phrase "baseline equivalence was confirmed", used in the earlier version of this manuscript, has been removed. What the data show is a small undetected difference favouring the control class, and the covariance model adjusts for it.

Confirmatory Result 1: Learning Engagement

The experimental class reported higher engagement than the control class by 10.92 points on the 15 to 60 scale, 95% CI [7.90, 13.94], $t(56) = 7.25$, $p < .001$, Holm-adjusted $p < .001$, Hedges' $g = 1.88$, 95% CI [1.26, 2.50]. Welch's test gave the same conclusion, $t(54.7) = 7.23$, as did the rank-based comparison, $U = 769$, $p < .001$. The advantage appeared in all three dimensions (Table 6 and Figure 2), and each dimension remains significant after Holm correction across the three. The dimension estimates should not be over-read, however. The cognitive difference is the largest at 4.05 points, 95% CI [2.94, 5.16], but its interval overlaps substantially with those for emotional, 3.50 points, 95% CI [2.45, 4.55], and behavioural engagement, 3.37 points, 95% CI [2.16, 4.58]. The ordering predicted in the theoretical framework, in which the cognitive component moves most, is consistent with the point estimates but is not established by them; on these data the three dimensions cannot be reliably ranked.

Table 6. Engagement by dimension, with confidence intervals

Dimension	Experimental *M* (*SD*)	Control *M* (*SD*)	Difference [95% CI]	$t(56)$	Holm *p*	Hedges' *g* [95% CI]
Emotional	16.50 (2.03)	13.00 (1.94)	3.50 [2.45, 4.55]	6.70	< .001	1.74 [1.14, 2.34]
Cognitive	16.37 (1.97)	12.32 (2.26)	4.05 [2.94, 5.16]	7.29	< .001	1.89 [1.27, 2.51]
Behavioural	15.73 (2.03)	12.36 (2.54)	3.37 [2.16, 4.58]	5.60	< .001	1.45 [0.88, 2.03]
Total	**48.60 (5.49)**	**37.68 (5.98)**	**10.92 [7.90, 13.94]**	**7.25**	**< .001**	**1.88 [1.26, 2.50]**

Note. Holm correction applied across the three dimensions. Dimension scores range from 5 to 20 and the total from 15 to 60. The confidence intervals for the three dimensions overlap substantially, so the dimensions cannot be reliably ranked on these data.

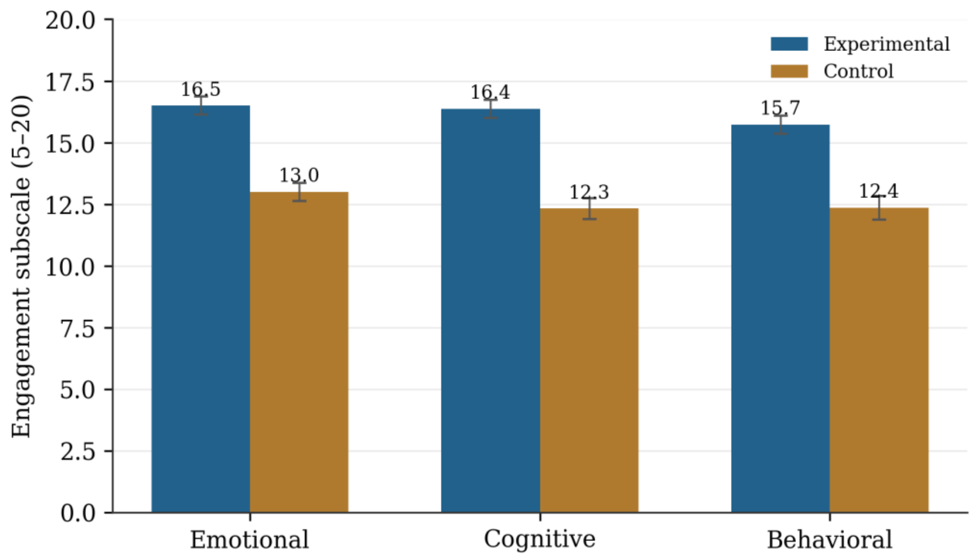

Figure 2. Mean engagement scores by dimension and group (error bars = standard error).

Confirmatory Result 2: Mathematics Achievement

For achievement the confirmatory test was the covariance analysis. Adjusted post-test means were 76.51 in the experimental class and 63.34 in the control class, an adjusted difference of 13.16 points, 95% CI [9.66, 16.67], $F(1, 55) = 56.73$, $p < 0.001$, Holm-adjusted $p < 0.001$, partial $\eta^2 = 0.51$, with a pooled within-group slope of 0.250 on the covariate. The adjusted estimate is slightly larger than the unadjusted difference of 12.88 points, which follows from the control class having started marginally ahead: removing that small head start moves the estimate in the experimental class's favour rather than against it.

Normalized gain is reported as a description of the same change rather than as further evidence for it. The experimental class recorded a mean $\langle g \rangle$ of 0.49 ($SD = 0.18$), in the medium band, against 0.18 ($SD = 0.18$) in the control class, in the low band (Hake, 1998), and Figure 3 shows the distributions. This measure depends on how far each student was from the ceiling at pre-test and its control-group distribution departed mildly from normality, so it is presented alongside the adjusted estimate rather than as a corroboration of it.

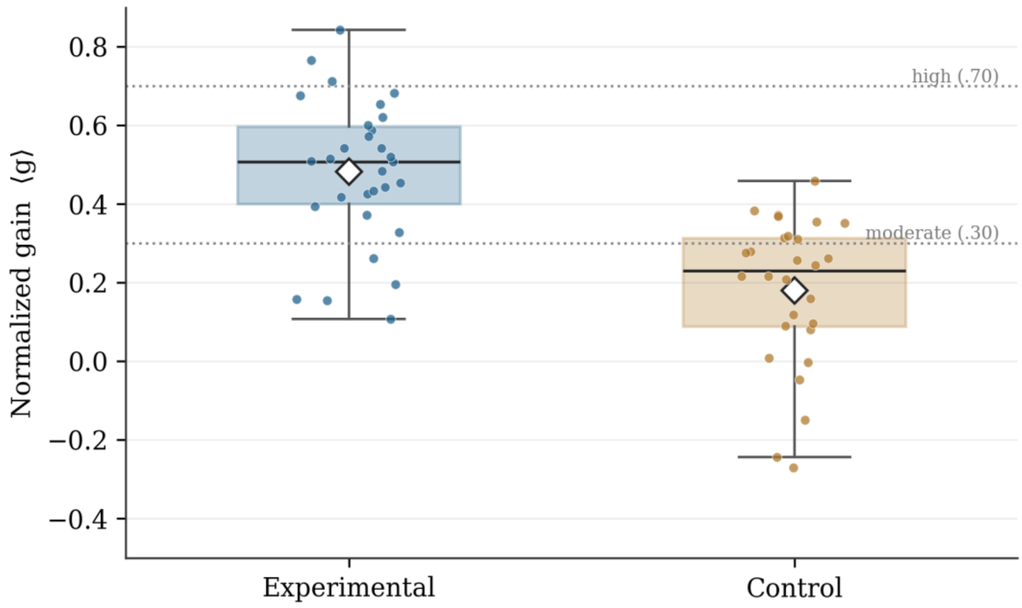

Figure 3. Distribution of normalized learning gains by group. Diamonds mark group means; dotted lines mark conventional moderate (0.30) and high (0.70) thresholds.

Exploratory Analysis: How engagement and achievement are related

Engagement and post-test achievement were strongly associated across the whole sample, $r = 0.74$, 95% CI [0.60, 0.84], $p < 0.001$. That pooled figure is partly an artefact of the design, however, because it combines two classes that differ on both variables, so the separation between the class means contributes to it. Within the classes the association is weaker, at $r = 0.43$ in the experimental class and $r = 0.62$ in the control class. Both remain positive and substantial, but the pooled correlation should not be quoted on its own.

When achievement was regressed on condition, engagement, and the pre-test together, all three predictors were significant: condition, $b = 6.61$, 95% CI [2.42, 10.80], $p = 0.003$; engagement, $b = 0.60$, 95% CI [0.33, 0.87], $p < 0.001$; and the pre-test, $b = 0.25$, 95% CI [0.05, 0.45], $p = 0.016$, $F(3, 54) = 33.45$, and $R^2 = 0.65$ (Table 7). The corresponding decomposition of the adjusted condition difference gives a path through engagement of 6.55 points, 95% percentile bootstrap CI [3.81, 9.61] from 10,000 resamples, and a remaining condition coefficient of 6.61 points, that is, approximately half of the adjusted difference on each side (Figure 5).

Table 5. Exploratory regression predicting post-test mathematics achievement

Predictor	*B* [95% CI]	β	*t*	*p*
Constant	27.32 [12.31, 42.34]	-	3.65	< 0.001
Condition (video-assisted DI)	6.61 [2.42, 10.80]	0.36	3.16	0.003
Engagement	0.60 [0.33, 0.87]	0.51	4.51	< 0.001
Pre-test achievement	0.25 [0.05, 0.45]	0.20	2.48	0.016

Note. R^2 = .65, adjusted R^2 = .63, $F(3, 54)$ = 33.45, p < .001.

What that decomposition means requires care, and the earlier version of this manuscript stated it too strongly. The bootstrap interval excludes zero, so the product term is not attributable to sampling noise. But engagement was measured once, after instruction, on the same occasion as the post-test, so nothing in the data establishes that the engagement difference preceded the achievement difference. A student who has just answered a test confidently may report having been engaged for that reason, which would produce the same numbers with the arrow reversed. Any unmeasured variable that raises both reported engagement and test performance, mathematics self-concept being the obvious candidate, would also produce them. The figure of roughly half should therefore be read as the share of the adjusted group difference that coincides statistically with the engagement difference in this sample, not as the share of an instructional effect that engagement carries.

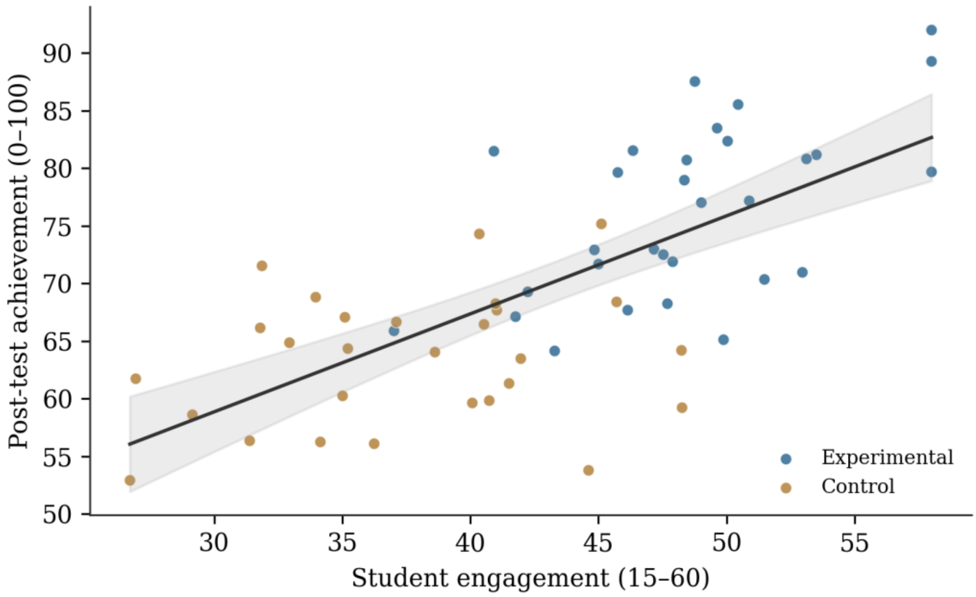

Figure 4. Relationship between student engagement and post-test achievement, with fitted regression line and 95% confidence band.

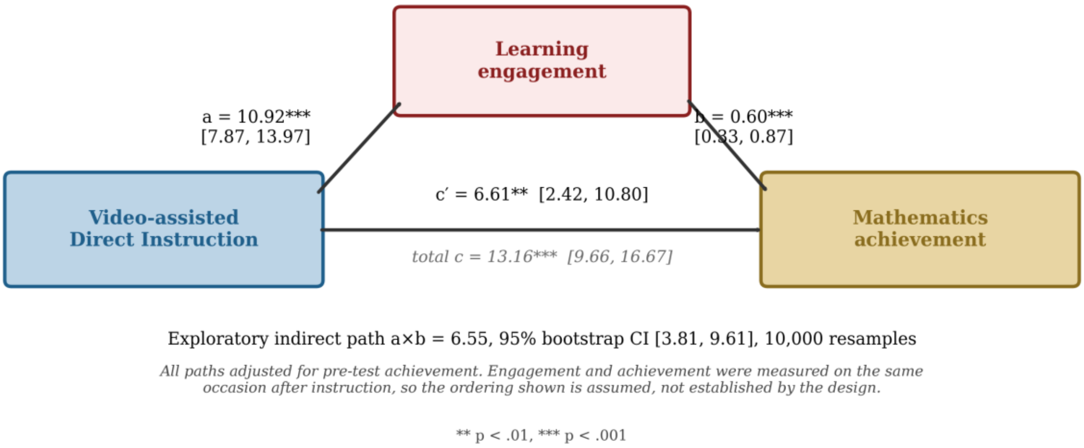

Figure 5. Path model relating video-assisted Direct Instruction, learning engagement, and mathematics achievement.

Discussion

Two classes taught the same content, for the same number of meetings, by the same teacher and following the same five-phase instructional sequence, ended the unit in different places. The class whose demonstration phase was carried by purpose-built video reported markedly higher engagement and, after adjustment for a small baseline difference that favoured the other class, scored about 13 points higher on the achievement post-test. That is the finding. The sections below consider what could produce it, what the engagement result can and cannot support, and under what conditions the result might transfer.

The most economical explanation runs through cognitive load rather than through motivation. In the control class the demonstration was transient: a step explained at the board and then erased is unavailable to a student who lost the thread, and the pacing belonged to the teacher. In the experimental class the same demonstration was segmented, re-playable, and paused between steps, which is precisely the format that the multimedia and cognitive-load accounts predict will free working-memory capacity for schema construction (Fyfield et al., 2022; Surur et al., 2023). On this account the videos worked as thinking aids that externalised steps a student would otherwise have to hold unaided (Zhu & Kaiser, 2022), and engagement is a correlate of that improved processing rather than its cause.

A second explanation attributes the difference to the clarity and consistency of the worked examples rather than to the medium as such. The recorded demonstrations were scripted, audited against a design specification, and identical across the six meetings, whereas a live exposition varies with the day. Some of the advantage may therefore reflect a better and more consistent worked example rather than video delivery, and a class receiving the same scripted exposition live might close part of the gap. A third explanation is time on task and teacher deployment: while the video ran, the teacher was free to circulate, which changes what happens during the demonstration phase even when its duration is held constant.

A fourth explanation is novelty. The experimental lessons took place in the computer laboratory with an unfamiliar arrangement, and attention responds to novelty in ways that decay. This is the standing objection to media-comparison studies, namely that the medium is confounded with the method, the materials, and the circumstances of their introduction, so an advantage for the media condition does not license a conclusion about the medium (Imaran et al., 2019). The present design cannot separate these four explanations. What it can do is make them visible, and the video specification in Table 4 at least makes the materials reproducible enough for a later study to hold them constant while varying the delivery.

The engagement difference is large and consistent across the three dimensions, and it is consistent with a literature holding that video sustains attention, makes abstract content concrete, and permits learner control of pace (Fyfield et al., 2022; Surur et al., 2023). The theoretical framework predicted that cognitive engagement would move most if the medium were doing pedagogical rather than affective work, and the point estimates are in that order. The intervals overlap, however, and with 58 students the three dimensions cannot be reliably ranked. The prediction is therefore consistent with the data rather than confirmed by them, and a study designed to test the ordering would need a substantially larger sample and repeated measurement of each dimension.

The relation between engagement and achievement is where the strongest claims in the earlier version of this manuscript have been withdrawn. Three features of the design bear on it. The questionnaire was administered once, after instruction, on the same occasion as the post-test, so there is no temporal separation between the putative mediator and the outcome and the direction of the association is not identified (Bullock-Yowell & Reardon, 2024). Both measures were collected from the same students at the same sitting, so common-method and mood variance are plausible contributors to the correlation (Rohanai et al., 2019). And the pooled correlation of .74, which the earlier version quoted, is inflated by the separation between the two class means; within the classes it falls to .43 and .62. The bootstrap decomposition, which places roughly half of the adjusted difference on the engagement path, is reported because it describes the data accurately, but it is a statistical partition of a group difference and not a measure of how much of an instructional effect engagement carries. Reading it as the latter would attribute to the design an identification it does not have.

This matters for what a teacher should take from the study. If engagement is on the causal path, then raising participation is a lever. If it is a concurrent indicator of better processing, then targeting participation directly may achieve little, and the operative change is the format of the demonstration. The present data cannot distinguish these, and the practical recommendation is therefore about the format rather than about engagement: use segmented, signalled, re-playable video to carry the demonstration phase, and treat the engagement gain as a welcome accompaniment whose causal status is unresolved.

The intervention required no new pedagogy and no equipment beyond the computer laboratories and internet already present at the site (Demszky et al., 2023; Reddy et al., 2021), which makes it a realistic option for schools with comparable infrastructure. Three costs should be stated plainly nonetheless. Producing six audited videos to the specification in Table 4 took substantially longer than preparing an equivalent series of board expositions, and that preparation is part of the intervention rather than an overhead to be discounted. Laboratory access for a full class is not universal in Indonesian vocational schools, and a version of this lesson delivered on shared or personal devices would not be the same intervention. And the effect estimated here belongs to video built to a documented specification, not to video in general; the review evidence indicates that design features and the addition-versus-substitution distinction moderate the benefit substantially (Noetel et al., 2021), so unspecified video should not be expected to reproduce these results.

The limitations bound the claims more tightly than the size of the estimates might suggest, and they are stated together here. Allocation was at class level by a coin toss between two intact classes, so unmeasured class-level differences cannot be excluded, and demographic covariates beyond programme and prior achievement were not recorded. The study took place in one school over six meetings on a bounded set of topics, so durability and transfer are untested and no delayed post-test was administered. The first author designed the intervention and taught both classes; exposure was equalised, plans were scripted and reviewed, fidelity was observed, and scoring was anonymised, but a teacher cannot be blinded to a condition he or she is delivering. Engagement rests on self-report with no observational or trace corroboration, was measured only once and concurrently with the outcome, and cannot therefore support a mechanism claim. Achievement was assessed with multiple-choice items, leaving problem-solving and transfer unmeasured. The experimental lessons took place in a different room from the control lessons, so setting is confounded with condition. And the effect sizes are far above what large-scale educational interventions typically produce, which in a single-site study with these features is a reason for caution rather than for confidence.

The next steps follow from those limitations rather than from the findings. The single most informative addition would be to measure engagement at least twice, once before and once during instruction, so that a change in engagement could be shown to precede the change in achievement; without that, no analysis of these data can move beyond association. A second would be to add observational coding or platform trace data so that reported participation can be checked against observed participation. A third would be a design that separates the candidate explanations, for example by comparing scripted video delivery against the same script delivered live, which would hold the worked example constant while varying the medium, and by extending the intervention long enough for novelty to decay. Multi-site replication with teachers who did not design the materials would test whether the effect survives the removal of expectancy, and achievement measures reaching into reasoning and transfer would show whether the advantage extends beyond recall.

Conclusion

Direct Instruction in which the presentation and demonstration phase was carried by purpose-built learning video produced higher reported engagement and higher mathematics achievement than Direct Instruction delivered by teacher exposition among tenth-grade vocational students. Engagement was higher by 10.92 points, 95% CI [7.90, 13.94], and adjusted post-test achievement by 13.16 points, 95% CI [9.66, 16.67], with the achievement estimate stable across separate-slopes, unadjusted, Welch, and rank-based analyses. Engagement and achievement were positively associated, and about half of the adjusted group difference coincided statistically with the engagement difference. That decomposition describes the data; it does not establish that engagement carried the effect, because engagement was measured once and on the same occasion as the outcome.

The practical recommendation therefore concerns the format of the demonstration rather than engagement as a target in its own right. Video built to an explicit specification, segmented into short learner-paced units, stripped of decorative material, signalling each step of a worked example, and narrated by the teacher who will lead the discussion around it, offers a low-cost way to change the phase of Direct Instruction where its known weakness lies. Whether that is what produced the difference observed here cannot be settled by this design: clarity of the scripted worked examples, the teacher's freedom to circulate while the video ran, the change of room, and the novelty of the arrangement are all live alternatives. These results were obtained in one school, over six meetings, by a teacher who designed the materials, and they are unusually large for an educational intervention. They should be read as a strong local result that identifies a promising and reproducible format, and that awaits replication with temporally ordered engagement measurement, observational corroboration, and teachers other than the designer.

Acknowledgment

The authors thank the principal, teachers, and tenth-grade students of the participating school for their cooperation, the colleague who observed every meeting and completed the fidelity and viewing-compliance checklists, and the lecturers of the Master Program in Educational Technology, Graduate School, Universitas Negeri Padang, for their guidance. The authors also thank the two anonymous reviewers, whose comments led to a substantial restructuring of the analysis, the reporting of the intervention, and the strength of the claims made. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

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

Bullock-Yowell, Emily., & Reardon, R. C. . (2024). Holland's RIASEC hexagon : a paradigm for life and work decisions. Florida State Open Publishing.

Chan, C.-C. (2018). Factors Affecting Career Goals of Taiwanese College Athletes From Perspective of Social Cognitive Career Theory. Journal of Career Development, 47(2), 193–206. https://doi.org/10.1177/0894845318793234

Christensen, R., Knezek, G., & Tyler-Wood, T. (2015). Alignment of Hands-on STEM Engagement Activities With Positive STEM Dispositions in Secondary School Students. Journal of Science Education and Technology, 24(6), 898–909. https://doi.org/10.1007/s10956-015-9572-6

Chukwuedo, S. O., Onwusuru, I. M., & Agbo, N. M. (2022). Practitioners' vocational guidance with direct learning model: influencing career commitment and employability in electrical/electronic technology education. International Journal for Educational and Vocational Guidance, 22(1), 23–48.

Creswell, J. W. , & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. . Sage Publications.

Demszky, D., Liu, J., Hill, H. C., & Jurafsky, D. (2023). Automated feedback improves teachers' uptake of student ideas in classroom discourse. Educational Evaluation and Policy Analysis, 45(3), 421–448. https://doi.org/10.3102/01623737231169270

Dewi, I. P., Asnur, L., Marta, R., Yanto, D. T. P., Dhanil, M., Saari, E. M., & Ali, S. K. (2025). How Effective Is Immersive AR Continental Food Course for Vocational Education? Analyzing Knowledge Gains and Learning Outcome Effects. International Journal of Information and Education Technology, 15(1), 127–136. https://doi.org/10.18178/ijiet.2025.15.1.2225

Doabler, C. T., Clarke, B., Kosty, D., Turtura, J. E., Sutherland, M., Maddox, S. A., & Smolkowski, K. (2021). Using direct observation to document "practice-based evidence" of evidence-based mathematics instruction. Journal of Learning Disabilities, 54(1), 20–35.

Entwistle, J., Smith, R., & Moore, T. (2025). Cognitive coaching and teacher efficacy: Evidence from school-based mentoring. Professional Development in Education, 51(2), 233–248. https://doi.org/10.1080/17521882.2025.2570687

Feng, J., & Hou, H. (2023). STEM in Vocational Education and Training: The Future Direction. Stem Education Review, 1. https://doi.org/10.54844/stemer.2023.0431

Fyfield, M., Henderson, M., & Phillips, M. (2022). Improving instructional video design: A systematic review. Australasian Journal of Educational Technology, 38(3), 155–183.

Guo, J., Zhou, Q., Xiong, S., Liu, H., Xia, S., & Su, R. (2023). Electric Power Training Simulation Platform Based on Virtual Reality Technology. 27. https://doi.org/10.1117/12.2691838

Guzey, S. S., Moore, T. J., & Harwell, M. R. (2016). Building Up STEM: An Analysis of Teacher-Developed Engineering Design-Based STEM Integration Curricular Materials. Journal of Pre-College Engineering Education Research (J-Peer), 6(1). https://doi.org/10.7771/2157-9288.1129

Imaran, M., Salisu, I., Aslam, H. D., Iqbal, J., & Irfan Hameed. (2019). Sustainability in the Era of IR 4 . 0 : The Mediating and. Processes, 7, 1–25.

Irfan, D., Mursyida, L., & Mubai, A. (2023). Implementation of mobile learning design in the flipped direct instruction model to increase student competency using a constructivist approach. Journal of Education Technology, 7(4), 752–762.

Irfan, D., Watrianthos, R., & Nur Bin Yunus, F. A. (2025). AI in Education: A Decade of Global Research Trends and Future Directions. International Journal of Modern Education and Computer Science, 17(2), 135–153. https://doi.org/10.5815/ijmecs.2025.02.07

Khan, D., Plopski, A., Fujimoto, Y., Kanbara, M., Jabeen, G., Zhang, Y. J., Zhang, X., & Kato, H. (2022). Surface Remeshing: A Systematic Literature Review of Methods and Research Directions. In IEEE Transactions on Visualization and Computer Graphics (Vol. 28, Number 3, pp. 1680–1713). IEEE Computer Society. https://doi.org/10.1109/TVCG.2020.3016645

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.

Krishnawati, N., Nurihsan, J., Budimansyah, D., & Nurdin, E. S. (2023). The Role of Entrepreneurship Education in Shaping Students' Emotional and Cognitive Competencies for Entrepreneurship. International Journal of Learning, Teaching and Educational Research, 22(2), 262–280. https://doi.org/10.26803/ijlter.22.2.15

Labak, S., Hursen, C., & Ozcinar, Z. (2025). Online video-based professional development for teachers: Effects on teaching practices and student learning. Education Sciences, 16(1), 36. https://doi.org/10.3390/educsci16010036

Lee, H., Cheng, Y., Wang, W., Lin, C., & Huang, Y. (2023). Exploring the Learning Process and Effectiveness of STEM Education via Learning Behavior Analysis and the Interactive-Constructive- Active-Passive Framework. Journal of Educational Computing Research, 61(5), 951–976. https://doi.org/10.1177/07356331221136888

Lutz, S., Ebenbeck, N., & Gebhardt, M. (2023). Mathematical Skills of Students with Special Educational Needs in the Area of Learning (SEN-L) in Pre-Vocational Programs in Germany. International Journal for Research in Vocational Education and Training, 10(1), 1–23.

Mayes, R. L., & Rittschof, K. A. (2021). Development of Interdisciplinary STEM Impact Measures of Student Attitudes and Reasoning. Frontiers in Education, 6. https://doi.org/10.3389/feduc.2021.631684

McLeod, S. (2018). Jean Piaget's Theory of Cognitive Development.

Mubai, A., Ambiyar, Irfan, D., & Rasul, M. S. (2023). Flipped Direct Instruction (FDI): A New Practicum Learning Model in Vocational Education. International Journal of Learning, Teaching and Educational Research, 22(7), 547–565. https://doi.org/10.26803/ijlter.22.7.29

Muskhir, M., Luthfi, A., Sidiq, H., & Fadillah, R. (2024). Development of Augmented Reality Based Interactive Learning Media on Electric Motor Installation Subjects. International Journal on Informatics Visualization, 8(4), 2097–2103. www.joiv.org/index.php/joiv

Noetel, M., Griffith, S., & Delaney, O. (2021). Video improves learning in higher education: A systematic review. Review of Educational Research, 91(2), 204–236. https://doi.org/10.3102/0034654321990713

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.

Pardimin, P., ARCANA, N., & Supriadi, D. (2019). Developing Media Based on the Information and Communications Technology to Improve the Effectiveness of the Direct Instruction Method in Mathematics Learning. Journal for the Education of Gifted Young Scientists, 7(4), 1311–1323. https://doi.org/10.17478/jegys.592636

Powell, S. R., Mason, E. N., Bos, S. E., Hirt, S., Ketterlin-Geller, L. R., & Lembke, E. S. (2021). A systematic review of mathematics interventions for middle–school students experiencing mathematics difficulty. Learning Disabilities Research & Practice, 36(4), 295–329.
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.
Rahmiati, R., Jalinus, N., Effendi, H., Fadillah, R., & Wulansari, R. E. (2024). Assessing the Impact of a STEM Learning Project Model on Artificial Intelligence Education in Higher Learning Institutions. Data and Metadata, 3. https://doi.org/10.56294/dm2024.623
Reddy, L. A., Dudek, C. M., & Lekwa, A. J. (2021). Coaching teachers for classroom behavior management: A randomized controlled trial. Journal of School Psychology, 87, 98–112. https://doi.org/10.1016/j.jsp.2021.04.001
Rohanai, R., Othman, H., Daud, K. A. M., Omar, N. H., Ahmad, M., Ismail, M. E., & Sulaiman, A. (2019). Concept of Correlation Between Active Learning and Employability Skills in TVET. Online Journal for Tvet Practitioners, 04(02). https://doi.org/10.30880/jtet.2020.05.01.003
Romero, I., & García, M. del M. (2023). Mathematical Attitudes Transformation When Introducing GeoGebra in the Secondary Classroom. Education and Information Technologies, 29(8), 10277–10302. https://doi.org/10.1007/s10639-023-12085-w
Santiago, L. M., & Nabayra, J. N. (2024). Video Assisted Learning Module (VALM) for the Least Learned Competencies in General Mathematics. MIER Journal of Educational Studies Trends and Practices, 482–510.
Satnarine, T. (2023). Systematic review methodology: Conducting high-quality reviews and understanding their significance in evidence-based practice. Journal for International Medical Graduates, 2.
Sezgin, S., & Yüzer, T. V. (2022). Analysing adaptive gamification design principles for online courses. Behaviour & Information Technology, 41. https://doi.org/10.1080/0144929X.2020.1817559
Shelle, G., Earnesty, D., Pilkenton, A., & Powell, E. (2018). Adaptive learning: An innovative method for online teaching and learning. The Journal of Extension. https://doi.org/10.34068/joe.56.05.17
Statti, A. (2020). Learner-Active, Technology-Infused Classroom. 38–56. https://doi.org/10.4018/978-1-5225-9304-1.ch003
Struyf, A., Loof, H. De, Pauw, J. B., & Petegem, P. Van. (2019). Students' Engagement in Different STEM Learning Environments: Integrated STEM Education as Promising Practice? International Journal of Science Education, 41(10), 1387–1407. https://doi.org/10.1080/09500693.2019.1607983
Sugiyono. (2016). Metode Penelitian Kuantitatif Kualitatif dan R & D. Alfabeta.
Suharno, S., Pambudi, N., & Harjanto, B. (2020). Vocational Education in Indonesia: History, Development, Opportunities, and Challenges. Children and Youth Services Review, 115, 105092. https://doi.org/10.1016/j.childyouth.2020.105092
Surur, A. M. S., Akhyar, M. K., Ridwanulloh, M. U., Mohamed, H. B., & Günerhan, H. (2023). Application of video-assisted online interactive learning to improve student mathematics learning outcomes. Jurnal Pengembangan Pembelajaran Matematika (JPPM), 5(2), 130–144.
Wannapiroon, P., Nilsook, P., Techakosit, S., & Kamkhuntod, S. (2021). STEM Literacy of Students in Vocational Education. International Journal of Technology in Education and Science, 5(4), 527–549. https://doi.org/10.46328/ijtes.253
Widana, I. W., Sopandi, A. T., & Suwardika, G. (2021). Development of an Authentic Assessment Model in Mathematics Learning: A Science, Technology, Engineering, and Mathematics (STEM) Approach. Indonesian Research Journal in Education |Irje|, 5(1), 192–209. https://doi.org/10.22437/irje.v5i1.12992
Zhou, X., Teng, D., & Al-Samarraie, H. (2024). The mediating role of generative AI self-regulation on students' critical thinking and problem-solving. Education Sciences, 14(12), 1302.
Zhu, Y., & Kaiser, G. (2022). Impacts of classroom teaching practices on students' mathematics learning interest, mathematics self-efficacy and mathematics test achievements: a secondary analysis of Shanghai data from the international video study Global Teaching InSights. ZDM–Mathematics Education, 54(3), 581–593.