From Participation to Legal Literacy: Causal Effects of Low-Cost Classroom Micro-Interventions in Universities

YANG Danming , LUO Haowen , LV Tong, WU Jinhe*, WU Miaoxian

Guangzhou College of Commerce, Guangzhou, Guangdong 510000, China

Abstract: This study tests whether embedding low-cost, replicable classroom “micro-interventions” that raise participation can improve university students’ legal literacy. We conducted a two-wave, class-level staggered adoption across law-related courses, assembling a four-wave student panel (T0–T3; N=1,080; 4,320 student-wave records). Causal identification relies on difference-in-differences with student and time fixed effects and class-clustered inference, complemented by an event-study; a 0–100 engagement index from learning logs opens the mechanism. The intervention increases the legal-cognition test (O1, 0–100) by 5.37 points (SE=0.33; d≈0.77) and the scenario-based compliance task (O2, 0–100) by 6.07 points (SE=0.58; d≈0.64). Dynamics show O1 building up across post-adoption waves, whereas O2 peaks early and attenuates. Benefits are larger for students with low baseline engagement, indicating compensatory effects. While mild positive pre-trends—especially for O2—advise caution in causal interpretation, robustness checks preserve sign and magnitude. Overall, routine, repeatable participatory practices embedded in ordinary teaching yield policy-relevant gains in legal cognition and short-term scenario performance and offer a scalable pathway for civics/legal education in higher education.

Keywords: Learning participation; Legal literacy; Active learning; Difference-in-differences; Event study; Higher education

I. Introduction

Legal literacy—the knowledge, skills, and dispositions that enable citizens to understand rules, evaluate claims, and act in accordance with the law—is an essential outcome of contemporary higher education and campus governance. A long tradition in learning sciences suggests that active participation—students’ moment-to-moment involvement in questioning, retrieval, explanation, and peer interaction—can deepen conceptual understanding and support transfer to authentic scenarios. In law-related courses specifically, student-active lectures and authentic assessment practices are increasingly recommended to connect doctrine with real-world decision-making. Yet universities still lack low-cost, routine classroom practices that reliably lift participation and deliver measurable gains in legal literacy at scale. Much of the existing evidence remains correlational or based on one-shot activities, making it difficult to separate selection into participatory classrooms from the causal impact of participation itself, or to trace how effects evolve over time.

* *Corresponding author: WU Jinhe*

This study addresses that gap by embedding a bundle of replicable, low-cost classroom "micro-interventions" into ordinary law-related courses and estimating their causal effects on students' legal literacy. The bundle—the "interactive ten"—comprises rapid roll-call polling, random cold/hot calling, one-minute papers, peer-defense mini-debates, a rolling key-point board/Pad wall, instant error-correction cards, authentic materials (e.g., school regulations, internship agreements, platform terms, online-infringement cases), brief role-plays, micro-assignments with immediate feedback, and a "minimal assignment + iterative versioning" principle. These routines operationalize mechanisms emphasized by learning science—attention/time-on-task, retrieval-based elaboration, timely feedback, and authenticity-driven relevance. We implement the bundle via staggered adoption at the class level across a semester and follow a four-wave student panel (N = 1,080) with two primary outcomes: an objective legal-cognition test (O1) and a scenario-based compliance task (O2). A process-oriented engagement index (0–100) built from learning-analytics logs opens the mechanism.

A. Design in Brief

Because adoption is rolled out at different times across classes and is one-way, we estimate effects with difference-in-differences models including student and time fixed effects and trace event-time dynamics to test pre-trends and post-adoption trajectories. This approach isolates within-student changes as their class transitions from pre- to post-adoption while using not-yet-treated classes as concurrent controls.

B. Preview of Findings (Non-Numerical)

Embedding these micro-interventions is associated with meaningful improvements in both legal cognition and scenario-based performance. Consistent with mechanism-based expectations, effects on conceptual knowledge (O1) build up across post-adoption waves, whereas effects on scenario transfer (O2) appear earlier but are less durable unless supported by sustained, context-rich practice. Benefits are larger for initially low-engagement students, indicating compensatory potential. These patterns align with theories of attention, retrieval, feedback, and authenticity in learning.

C. Contributions

(1) We provide causal evidence on the participation-to-literacy pathway using a class-level staggered design with event-time dynamics, advancing beyond cross-sectional associations and one-shot classroom experiments.(2) We specify a replicable set of low-cost micro-interventions—the interactive ten—that instructors can embed into routine teaching without special hardware or additional class time, offering a pragmatic lever for civics/legal education.(3) We foreground open and reproducible practice by documenting variables and standardized effect translations, and we examine heterogeneity with an equity lens (e.g., baseline engagement, major, gender).

D. Roadmap

Section 2 situates the study within research on participation, active learning, and legal/civic literacy and lays out the conceptual framework linking participation to cognition and transfer. Section 3 details the context, intervention, measures, and ethics. Section 4 presents identification and estimation—difference-in-differences and event-study—alongside diagnostics and robustness. Section 5 reports main effects, dynamics, heterogeneity, and process evidence. Section 6 discusses implications for scalable classroom practice and governance-oriented education, as well as limitations and directions for future work.

II. Related Work and Conceptual Framework

A. Positioning in the Literature

A large body of work in education and the learning sciences argues that active participation—visible behaviors such as responding to questions, explaining ideas, practicing retrieval, and engaging with peers—improves conceptual mastery and transfer.

This study contributes to that literature by combining a staggered-adoption classroom design with event-study dynamics to estimate the causal effect of participation on two distinct learning targets: (i) legal-cognition—core concepts, principles, and rules; and (ii) scenario performance—choosing compliant strategies under realistic contractual, regulatory, or platform cases. The distinction mirrors the "near vs. far transfer" debate: gains in structured knowledge may accumulate with practice, whereas transfer to complex, ill-structured scenarios may require longer or different forms of practice and contextually rich feedback.

B. Construct Definitions and Measures

We conceptualize learning participation as a multi-component construct:

(1) Behavioral participation (doing): attendance, responding to polls, random cold/hot calling, submitting micro-assignments on time.

(2) Cognitive participation (thinking): one-minute papers (retrieval/elaboration), peer defense (generative explanation, perspective taking), error-correction cards (productive failure).

(3) Affective/relational participation (wanting/belonging): perceived relevance (authentic materials), teacher–student interaction quality, and social accountability.

These components map to our empirical indicators:

(4) Engagement index (0–100), constructed from logs of sign-ins, hand-raising/speaking, instant polling, timeliness of submission, Pad messages, and homework completion—our primary process mediator.

(5) Attitudinal scales (Cronbach's $\alpha \geqslant 0.80$ in prior work): active participation, teacher–student interaction, and content relevance to life.

(6) Outcomes: legal_cognition (O1, 0–100) and scenario_score (O2, 0–100).

The "interactive ten" micro-interventions used in this study operationalize these components at minimal cost: rapid polling and one-minute papers (retrieval and feedback); random calling and peer defense (accountability and elaboration); rolling key-point board/Pad wall (shared attention and externalization); correction cards (error-based learning); authentic materials and role-play (situational interest and transfer); micro-assignments with immediate feedback and a "minimal assignment + iteration" principle (spacing and iterative refinement).

C. Theory of Change and Mechanisms

We posit complementarity between interaction and relevance: authentic, life—proximal materials raise stakes and attention; interactive routines turn that attention into elaboration and retrieval. This implies a positive synergy term (Relev × Interact) on outcomes, especially for O2 where context and transfer matter.

$$D_{ct} \Rightarrow \underset{\text{\textbackslash scriptsize polling, calling}}{\text{Attention \textbackslash\& time-{}-on-{}-task}} + \underset{\text{\textbackslash scriptsize retrieval, peer defense, one-{}-minute papers}}{\text{Deep processing}} + \underset{\text{\textbackslash scriptsize instant correction, micro-{}-feedback}}{\text{Timely feedback}} + \underset{\text{\textbackslash scriptsize real cases, role-{}-play}}{\text{Relevance \textbackslash\& authenticity}}$$

$$\Rightarrow M_{ict}(\text{engagement, attitudes}) \Rightarrow Y_{ict}(\text{O1 legal cognition, O2 scenario performance})$$

The framework yields distinct dynamic predictions:

For O1 (legal cognition), repeated retrieval and feedback should produce monotonic growth post-adoption as knowledge consolidates (spacing + desirable difficulties).

For O2 (scenario transfer), effects may appear earlier but attenuate without extended, context-rich practice (e.g., role-play, authentic case cycles), because transfer requires abstraction and schema mapping beyond recall.

These predictions align with the event—study patterns observed in our data: O1 expands from roughly +5.8 to +8.0 points across k=0–2, whereas O2 drops from ≈+3.8 to ≈+0.5.

D. Theoretical Rationale for Methods and Hypotheses

Our conceptual model posits that routine, light-touch interactive routines increase moment-to-moment participation, which in turn enhances attention, retrieval-based elaboration, and timely feedback. These processes are expected to accumulate conceptual knowledge (legal cognition) while producing earlier but more fragile gains in scenario-transfer unless authentic, context-rich practice continues. This mechanism-based view yields dynamic predictions: monotonic post-adoption growth for legal cognition and early peaks with potential attenuation for scenario performance.

The staggered adoption at the class level creates quasi-experimental variation over time and across classes. Because adoption is one-way and scheduled rather than outcome-driven, a difference-in-differences framework with student and time fixed effects isolates within-student changes as their class transitions from pre- to post-adoption, using not-yet-treated classes as concurrent controls. An event-study specification is essential here: it directly tests the pre-trend requirement and traces dynamics that our theory predicts (build-up for conceptual knowledge; early, less durable transfer). References on identification and dynamics include Roth et al., de Chaisemartin & D'Haultfoeuille, and Rambachan & Roth.

Guided by this rationale, we pre-specify six hypotheses that map mechanisms to outcomes and dynamics (H1–H6). H1–H3 state average and dynamic effects for the two outcomes; H4–H6 articulate compensatory heterogeneity, dose–response, and fairness expectations. The empirical design and estimators reported below (Sections 3–4) follow directly from these theoretical commitments.

E. Formalization and Testable Hypotheses

Let denote participation intensity (proxied by the engagement index), and let be baseline covariates. In potential-outcomes notation,

$$Y_{ict} = f(D_{ct}, P_{ict}, D_{ct} \times P_{ict}, X_{ict}, \gamma_c, \delta_t, \varepsilon_{ict}),$$
$$P_{ict} = g(D_{ct}, X_{ict}, \gamma_c, \delta_t, u_{ict})$$

where γ_c and δ_t are class and time effects. The micro-interventions increase P_{ict} contemporaneously (first stage), and outcomes respond both directly to (e.g., through structured practice embedded in class) and indirectly through P_{ict} (mediation).

From this setup we derive six testable hypotheses:

H1 (Average effect). The average treatment effect on the treated (ATT) > 0 for both O1 and O2.

H2 (Dynamics—O1). The dynamic treatment effect is non-decreasing in event time ⩾ 0 over the first few post periods.

H3 (Dynamics—O2). The dynamic treatment effect peaks early and attenuates absent additional scenario practice.

H4 (Heterogeneity). Effects are larger for students with lower baseline participation, reflecting compensatory benefits of accountability and feedback.

H5 (Dose–response). Conditional on treatment, higher implementation intensity (frequency/

coverage of the "ten") yields larger but concave gains; the Relev × Interact synergy is positive, especially for O2.

H6 (Fairness). Residuals and improvements are calibrated across groups (gender, major, grade); if not, the framework predicts stronger gains where baseline engagement is structurally lower.

F. Empirical Implications for the Design

The framework guides our empirical choices in three ways.

Event-study focus. Because mechanisms predict distinct dynamics for O1 ersus O2, we estimate event-time coefficients that trace build-up and attenuation rather than relying on a single average effect.

Process measurement and mediation. The engagement index operationalizes P_{ict} and allows for mediation tests (indirect effect $D \rightarrow P \rightarrow Y$) and dose-response analysis.

Subgroup analysis and equity. Heterogeneity by baseline engagement/major/gender is not ancillary: it is a prediction of the theory (H4, H6). We therefore pre-specify subgroup contrasts and report standardized effects alongside point-gains.

In sum, this study integrates insights from active learning, retrieval-based practice, formative assessment, and authenticity-driven transfer into a coherent participation-to-literacy framework. The "interactive ten" are designed to activate the mediators that the framework highlights, while our staggered-adoption design and event-study estimation connect those mechanisms to causal effects and their temporal profiles.

III. Design and Data

A. Setting, Sample, and Timeline

We evaluate a bundle of participation-oriented micro-interventions in regular, credit-bearing law-related courses across multiple colleges. The analytic panel comprises 1,080 students nested in roughly two dozen classes taught by comparable instructor teams, observed over four waves (T0–T3), yielding 4,320 student-wave records. Classes adopt the intervention in two waves—A1 early in the semester and A2 later—creating quasi-experimental variation over time and across classes.

The analytic panel includes 1,080 students observed over four waves. At baseline, about 37% of students were law majors (≈ 400 students) and about 56% had previously completed at least one law-related course (≈ 605 students); values are rounded from the two adoption cohorts' baseline means. Baseline attendance was high (≈ 90%). These descriptive facts complement the balance checks reported below and clarify that our treated and not-yet-treated classes are comparable prior to adoption.Figure 1 visualizes the staggered adoption schedule across the four observation waves (T0–T3), with classes adopting the intervention in two waves—A1 early and A2 later in the semester.

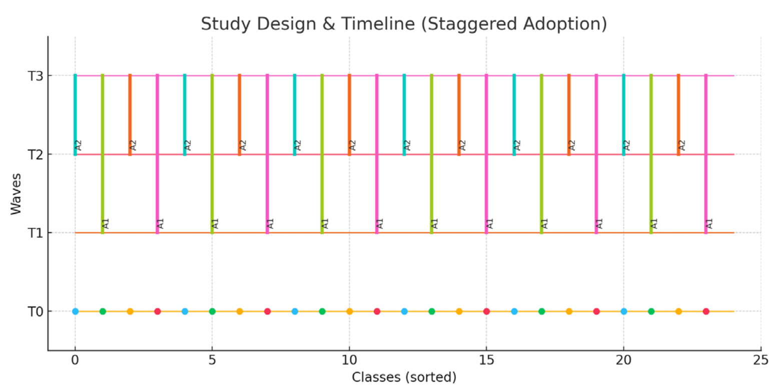

Figure 1. Study design and staggered adoption timeline at the class level.

B. Intervention: the "interactive ten"

The intervention is deliberately low-cost and replicable, consisting of ten micro-activities that instructors embed into ordinary lessons:

(1) rapid roll-call polling;

(2) cold/hot random calling;

(3) one-minute papers (muddiest point);

(4) peer-defense mini debates;

(5) rolling key-point board / Pad wall;

(6) instant error-correction cards;

(7) authentic materials (rules, contracts, platform terms, online infringement);

(8) brief role-plays;

(9) micro-assignments with immediate feedback;

(10) minimal assignment + iterative versioning.

These elements target attention/time-on-task, deep processing (retrieval and explanation), timely feedback, and relevance/authenticity (see Section 2).

C. Measures

We collect outcomes, process indicators, and covariates at the student-wave level. Scores are placed on a common 0–100 scale for interpretability.

Table 1 reports variables and measurement (student–wave level unless noted).

Table 1. Variables and measurement(student - wave level unless noted)

Construct / Variable	Definition and measurement	Scale
Identification & panel keys		
student_id, class_id, teacher_id, school_id	Unique identifiers for student, class (treatment unit), teacher, and school	—
wave	Time index: T0 baseline, T1, T2, T3	0 - 3
adoption_wave	Class adoption wave (A1 or A2)	1/2
treat_post	Treatment indicator: 1{ wave ≥ adoption_wave(class) }	0/1
rel_time (derived)	Event time: wave - adoption_wave(class); reference k=-1	integer
Primary outcomes		
legal_cognition (O1)	Objective test (30 - 40 items) covering legal foundations, campus governance, contracts/IP/online infringement; KR-20 ≥ .70	0 - 100
scenario_score (O2)	Scenario-based task with authentic materials; select compliant strategies; rubric-scored	0 - 100
Process / mediators		
engage_idx	Engagement index from logs: sign-ins, hand-raising/speaking, instant polling, on-time submission, Pad messages, homework completion (z-scores → weighted sum, rescaled)	0 - 100
att_active_participation, att_teach_interaction, att_close_to_life	3 Likert scales (prior work α ≥ .80; one-factor PCA >50% variance)	0 - 100
attendance_rate	Share of sessions attended within wave	0 - 100
Baseline covariates (T0 unless noted)		
gender, grade	Student gender; year/grade	binary/int
law_major, took_law_course	Law major; prior law course taken	0/1

Scoring and reliability. The O1 test pools items over legal foundations, campus rules, contracts/IP/online infringement; internal consistency is assessed with KR-20 (target ≥ .70). O2 rubric reliability is checked by double-rating a random subset and reporting agreement statistics (e.g., or ICC). Attitude scales use 4–5-point Likert items, aggregated and re-scaled to 0–100; internal consistency (Cronbach's) and a one-factor PCA check (>50% variance) are reported in the

Appendix.

Engagement index. Let be standardized logs (e.g., polling count, on-time submission). The index is

$$engage\_idx_{ict} = 100 \cdot \frac{\sum_m W_m z_{mict}}{\sum_m ||}$$

with pre-specified (equal weights in the main analysis; alternatives used in robustness).

D. Treatment Assignment and Event time

Treatment is assigned at the class level by administrative scheduling rather than by prior outcomes. For class at wave,

$$D_{ct} \equiv treat\_post_{ct} = 1\{t \geqslant adoption\_wave_c\},$$
$$rel\_time_{ct} = t - adoption\_wave_c$$

with $k = -1$ (the last pre-period) as the event-study reference period. Students exposed to treatment are those enrolled in treated classes; contemporaneous not-yet-treated classes serve as controls.

E. Pre-processing and Data Quality

We pre-register and implement the following steps:

(1) Rescaling outcomes and scales to 0–100;

(2) winsorization at 1–99% in robustness checks (main results use raw 0–100 bounds);

(3) panel integrity checks (four records per student; attrition accounting);

(4) coding of covariates at T0 to avoid "bad controls";

(5) construction of event-time dummies with $k \leqslant -2$ optionally collapsed into a single "far lead" bin;

(6) cluster identifiers verified for within-class correlation.

F. Baseline Balance

At T0, A1 and A2 classes are highly comparable. Table 2 reports group means and standardized mean differences (SMD) for outcomes, engagement, attitudes, and covariates; the maximum |SMD|≈0.085, comfortably below the 0.10 convention.

Table 2. Baseline (T0) balance by adoption cohort (A1 vs. A2)

Variable (T0)	A1 Mean	A2 Mean	SMD	Diff (A1-A2)	|SMD|	Flag
legal_cognition	52.49	52.87	−0.055	−0.38	0.055	Good(0.05 - 0.10)
scenario_score	53.62	53.86	−0.025	−0.24	0.025	Excellent(<0.05)
engage_idx	57.87	57.94	−0.005	−0.07	0.005	Excellent(<0.05)
attendance_rate	0.90	0.90	0.009	0.00	0.009	Excellent(<0.05)
gender	0.46	0.51	−0.085	−0.04	0.085	Good(0.05 - 0.10)
grade	2.45	2.44	0.009	0.01	0.009	Excellent(<0.05)
law_major	0.35	0.39	−0.081	−0.04	0.081	Good(0.05 - 0.10)
took_law_course	0.57	0.55	0.041	0.02	0.041	Excellent(<0.05)
att_active_participation	3.59	3.61	−0.035	−0.02	0.035	Excellent(<0.05)
att_teach_interaction	3.72	3.73	−0.015	−0.01	0.015	Excellent(<0.05)
att_close_to_life	3.79	3.78	0.009	0.01	0.009	Excellent(<0.05)

Notes. SMD = standardized mean difference (A1 vs. A2). Continuous variables scaled 0–100. Thresholds: <0.05 Excellent; 0.05–0.10 Good; 0.10–0.25 Needs attention; ⩾ 0.25 Large. Maximum |SMD| = 0.085.

G. Ethical Procedures and Data Governance

The study is minimal-risk classroom research. Students receive brief information sheets and provide informed consent; participation in learning activities is part of routine instruction, whereas data use is opt-out. Identifiers are hashed and stored separately from analysis files; only de-identified analytic datasets are shared. We follow institutional ethical guidelines for learning analytics, restrict access by role, and comply with applicable data-protection regulations. Any open release redacts small cells and indirect identifiers.

H. Power and Minimum Detectable Effects (MDE)

Given the observed variance, cluster structure (class-level treatment), and four measurement waves, the design detects policy-relevant effects. Using cluster-robust standard errors from the main DID:

O1 (legal cognition): (1.96 + 0.84)×0.335≈0.94 points.

O2 (scenario score): 1.63 points.

In reporting, we present both raw-point effects and standardized effects (Cohen's) relative to T0 standard deviations to facilitate cross-study comparison.

IV. Identification and Estimation

A. Design-based Identification and Assumptions

We exploit staggered adoption at the class level. Classes adopt the participation bundle in two waves (A1, A2); students are observed for four waves t ∈ {0,1,2,3}. Identification comes from within-students change before/after their class adopts, with not-yet-treated classes serving as concurrent controls.

We work under the following conditions (stated for class c, student i, wave t):

(1) Conditional parallel trends (relative to k $=-1$). Absent treatment, treated and not-yet-treated classes would have followed common trends relative to the last pre-period($k=-1$) after conditioning on fixed effects and baseline covariates. We allow far leads to deviate and, for inference, collapse far leads into a single bin ($k \leqslant -2$).

(2) Limited anticipation (last pre-period only). Potential outcomes at $k = -1$ are unaffected by future adoption; we do not impose no-anticipation for earlier periods ($k \leqslant -2$).

(3) Monotone treatment path. Adoption is one-way (once treated, always treated); no reversal.

SUTVA at the class level. Outcomes for student i in class c depend on Dct in their own class; cross-class spillovers are negligible. As a sensitivity, we allow for teacher×time or school×time fixed effects to absorb common shocks within instructors/schools.

As-good-as random timing (conditional). Adoption timing is organizational/scheduling-driven and uncorrelated with time-varying unobservables after controlling for fixed effects and baseline covariates.

Diagnostics. We diagnose Assumption 1 with an event-study that collapses $k \leqslant -2$ and reports a lead test for $\beta\_\{k=-2\}=0$; given the positive far lead for O2, we treat TWFE dynamics as descriptive and rely on Sun-Abraham and Callaway-Sant'Anna dynamic ATTs for robustness (see Sections 4.6 and 5.3).

B. Baseline DID (Two-way Fixed Effects)

Let be either O1 legal_cognition or O2 scenario_score (scaled 0–100).

Let be the class-level treatment indicator,

baseline covariates, class fixed effects, and time fixed effects.

Equation (TWFE specification):

$$Y_{ict} = \alpha + \beta D_{ct} + \gamma_c + \delta_t + \theta^{\top} X_{ict} + \varepsilon_{ict}.$$

is the average treatment effect on the treated (ATT) under Assumptions 1–5.

absorbs time-invariant class/teacher features; captures common shocks.

use T0 values (gender, grade, law_major, took_law_course, attendance) to improve precision without "bad controls".

Inference. We report class-clustered robust SEs (Arellano type). Given the finite number of clusters (≈ classes), we additionally compute wild-cluster bootstrap and CR2 (Bell–McCaffrey) corrections in robustness tables.

Effect translation. Alongside point gains, we report Cohen's using the T0 SD:

Minimum detectable effects (MDE) use the empirical.

C. Event Study (Dynamic Effects and Pre-trends)

Define event time and set (last pre-period) as the reference. We estimate:

$$Y_{ict} = \alpha + \sum_{k \in K \setminus \{-1\}} \beta_k 1\{t - adoption\_wave_c = k\} + \gamma_c + \delta_t + \theta^\top X_{ict} + \varepsilon_{ict}.$$

Pre-trend test. $H_0{:}\beta_k=0$ for all, $k \leqslant -2$ (joint Wald with class-clustered covariance).

Dynamic profile. trace build-up/attenuation after adoption. In your estimates, O1 increases from≈+5.8 (k=0) to≈+8.0 (k=2); O2 attenuates from≈+3.8 to≈+0.5.

Heterogeneous effects and TWFE weighting. Because TWFE can be biased under treatment-effect heterogeneity in staggered designs, we treat this as descriptive and report heterogeneity-robust event studies using Sun–Abraham or Callaway–Sant'Anna aggregations as robustness (Section 4.6).

SA/CS implementation (reporting, not code). We estimate group-time effects by adoption cohort and aggregate to dynamic ATT relative to the pre-period. These complement TWFE and are the preferred specification if pre-trends differ.

D. Weighting and Balancing

To mitigate selection into early vs. late adoption, we optionally re-weight observations with propensity-score or entropy-balancing weights constructed from class-level T0 covariates (e.g., class means of gender, grade, law_major, baseline O1, baseline engagement). Let be the stabilized weight.

Equation (weighted DID):

$$\hat{\beta}^{wDID} = argmin_\beta \sum_{i,c,t} w_{ict}(Y_{ict} - \alpha - \beta D_{ct} - \gamma_c - \delta_t - \theta^\top X_{ict})^2 eq: weighted$$

We report pre/post-weight balance via standardized mean differences (SMDs). Weighted DID and weighted event studies are presented as robustness.

E. Heterogeneity, Dose–response, and Mediation

Heterogeneity. For subgroup G ∈ {low baseline engagement, law major, gender},

$$Y_{ict} = \alpha + \beta D_{ct} + \textstyle\sum_g \beta_g (D_{ct} \times 1\{G = g\}) + \gamma_c + \delta_t + \theta^\top X_{ict} + \varepsilon_{ict}. eq: heter$$

We report ATEs for $G = 0/1$ as β and β_g. Your results indicate larger O1 gains among low-engagement students (interaction≈+0.95).

Dose-response. Where instructor logs allow, define intensity as the frequency/coverage of the "interactive ten" in class at. Estimate:

$$Y_{ict} = \alpha + \beta_1 D_{ct} + \beta_2 d_{ose} + \beta_3 (D_{ct} \times d_{ose}) + \gamma_c + \delta_t + \theta^\top X_{ict} + \varepsilon_{ict}. eq: dose$$

captures marginal gains per unit of implementation among treated classes.

Mediation (process). With engagement, M_{ict},

$$M_{ict} = \eta_0 + \eta_1 D_{ct} + \gamma_c + \delta_t + \kappa^\top X_{ict} + u_{ict}$$

$$Y_{ict} = \alpha + \tau D_{ct} + \delta M_{ict} + \gamma_c + \delta_t + \theta^\top X_{ict} + \varepsilon_{ict}.$$

The indirect effect is with class-clustered bootstrap CIs (mediator measured contemporaneously or with a lag for temporal ordering).

F. Robustness, Placebo, and Sensitivity Analyses

For robustness, we only retain interpretable, inference-worthy specifications: (1) student & time TWFE with class-clustered CR2/Arellano standard errors; and (2) Sun-Abraham / Callaway–Sant'Anna dynamic ATTs.

Estimates with Teacher×Time (Teacher×Wave) FE are moved to the Appendix and labeled "illustration; not for inference," because high-dimensional FE over-absorbs the treatment path and renders SEs numerically near zero. All inference in the main text relies on class-clustered CR2/Arellano and SA/CS results.

(1) Event-study leads (pre-trends). Joint Wald/F test of.

(2) Alternative FE and clustering.

Teacher×Time and/or School×Time FE (reported in Appendix as illustration; SEs numerically ~0 due to over-absorption — not for inference);

Class-specific linear trends;

Teacher-clustered SEs and multi-way clustering as sensitivity.3. Wild-cluster bootstrap p-values (Rademacher weights) for finite clusters.

(3) Sun–Abraham / Callaway–Sant'Anna dynamic ATT to address TWFE weighting under heterogeneity.

(4) Placebos. (i) Shuffled adoption: randomly permute adoption_wave_c across classes; (ii) pseudo-treatment: assign treatment in pre-period only; (iii) leave-one-teacher/class out.

(5) Winsorization/trim of outcomes at 1–99% to check outlier sensitivity.

(6) Missing data: (i) IPW for attrition using T0 covariates and lagged outcomes; (ii) MICE imputation in a sensitivity column.

We summarize directional stability (sign of $\hat{\beta}$), magnitude bands, and overlap of CIs across specifications (Table 3).

Table 3. Robustness across specifications

Outcome	Specification	Effect (β, points)	SE
O1 Legal cognition	A.Base (cluster=class)	5.37	0.33
O1 Legal cognition	B. Cluster=teacher	5.37	0.33
O1 Legal cognition	C. + Teacher×Time FE	5.48	≈0†
O1 Legal cognition	D. + Class trend	5.85	0.37
O1 Legal cognition	E. No covariates	5.37	0.33
O2 Scenario score	A. Base (cluster=class)	6.07	0.58
O2 Scenario score	B. Cluster=teacher	6.07	0.58
O2 Scenario score	C. + Teacher×Time FE	2.34	≈0†
O2 Scenario score	D. + Class trend	6.52	0.64
O2 Scenario score	E. No covariates	6.07	0.58

Notes. Class-clustered SEs unless otherwise noted. † SE numerically ~0 under teacher×time FE due to over-absorption/collinearity; do not use for inference.

We complement TWFE with group–time effects that are robust to heterogeneous treatment timing. We estimate dynamic ATTs by adoption cohort and aggregate them following Sun–Abraham or Callaway–Sant'Anna, using k = −1 as the reference and collapsing far leads to k≤−2. Inference uses class-clustered standard errors and wild-cluster bootstrap as a sensitivity check. Reporting includes: (i) the event-time plot with 95% CIs (Figures 3–4), (ii) a lead test for: all leads = 0

=0 (with $k \leqslant -2$ collapsed, this reduces to testing; Table 4), and (iii) post-adoption dynamics for $k \in \{0,1,2\}$ (Table 5). These diagnostics guide interpretation when pre-trend deviations are present, notably for O2.

Table 4. Lead test of pre-trends (H_0: all leads = 0; with $k \leqslant -2$ collapsed, this reduces to testing β_{k = −2} = 0). Class-clustered SEs.

Outcome	Lead (k = 2) beta	SE	z	p-value	95% CI
O1: legal_cognition	0.913	0.363	2.381	0.017265	(0.191, 1.635)
O2: scenario_score	3.6	0.587	6.132	0.0	(2.449, 4.751)

Table 5. Post-adoption dynamics relative to k = −1 (k ∈ {0, 1, 2}). Class-clustered SEs.

Outcome	k (post)	beta	SE	95% CI
O1: legal_cognition	0	5.793893637522067	0.24575616200567	(5.312, 6.276)
O1: legal_cognition	1	-6.898055977609031	0.246270864077544	(-7.381, -6.415)
O1: legal_cognition	2	-7.90701783505036	0.370198758978513	(-8.631, -7.183)
O2: scenario score	0	3.77175812926359	0.43812458012513	(2.913, 4.630)
O2: scenario score	1	2.168930730304892	0.365728513682205	(1.453, 2.885)
O2: scenario score	2	-1.248571057948046	0.539511835364873	(-2.307, -0.190)

G. Fairness and Calibration

We compare groupwise residuals and gain distributionsby gender/major/engagement strata and report calibration slope/intercept from regressing realized outcomes on fitted values. We also report interaction effects (Eq. (heter)) to identify systematic under- or over-estimation for subgroups.

H. Multiple testing and reporting standards

Primary outcomes are two (O1, O2). We control the family-wise error rate using Holm–Bonferroni for these two and report BH-FDR for exploratory contrasts (heterogeneity, mediation). All tables report: point estimate, class-clustered SE (and wild-cluster p as footnote), 95% CI, N, number of classes, fixed-effects structure, and whether weights are used.

V. Results

A. Descriptives and Baseline Comparability

At baseline (T0), A1 and A2 classes are highly comparable across outcomes, process indices, and covariates (Table 2). The maximum absolute SMD ≈ 0.085, well below conventional thresholds(0.10), supporting the design's credibility for difference-in-differences (DID).

B. Main Effects (DID with Student and Time Fixed Effects)

O1—Legal cognition. Treatment raises scores by 5.37 points (SE = 0.33; 95% CI [4.72, 6.03]; p <.001). Standardized, this is Cohen's d ≈ 0.77, which corresponds to a U3 percentile shift of ~78% (i.e., moving the median student to the 78th–79th percentile of the baseline distribution).

O2—Scenario performance. Treatment increases scores by 6.07 points (SE = 0.58; 95% CI [4.93, 7.21]; p <.001), d≈0.64 (U3 ≈ 74%).

Given the observed standard errors, the minimum detectable effect (80% power) is≈ 0.94 points for O1 and 1.63 points for O2, implying that the study is well powered for policy-relevant impacts. Holm–Bonferroni across the two primary outcomes leaves both results significant.

C. Dynamics and Pre-trend Diagnostics (Event Study)

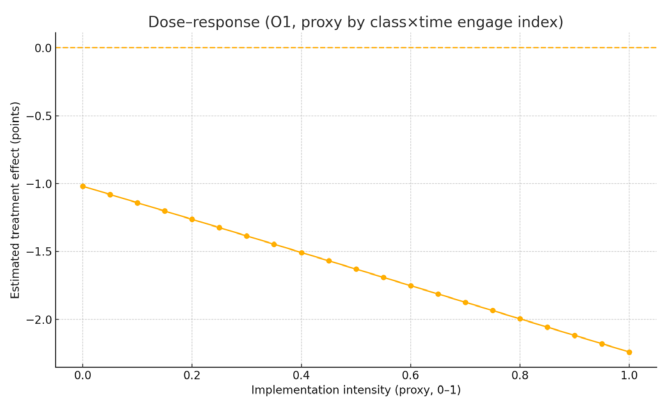

Figure 2. Event-study dynamic effects for O1 (legal cognition)

O1 displays monotonic post-adoption growth in the event-study (TWFE) coefficients:+7.97 [7.24, 8.69] (all 95% CIs exclude zero). The far-lead bin (k ⩽ −2) is small but positive (≈ +0.91, 95% CI [0.16, 1.66]), indicating a mild upward drift prior to adoption.

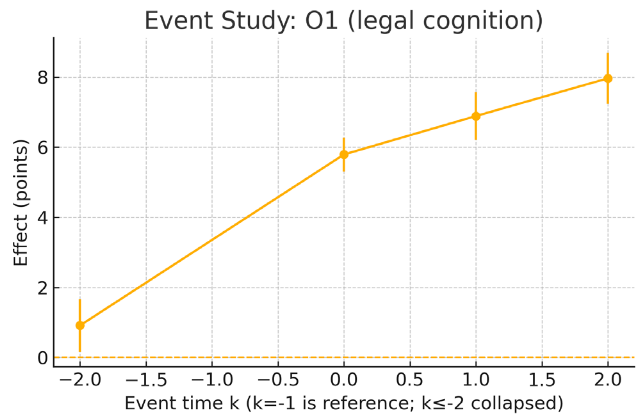

Figure 3. Event study (O1, legal cognition; TWFE coefficients with 95% CIs)

O2 starts positive and attenuates in the event-study (TWFE) coefficients:+0.47 [−0.65, 1.59] (the last includes zero). The far-lead bin (k≤−2) is clearly positive (≈ +3.60, 95% CI [2.45, 4.75]), which cautions against strong parallel-trend claims for O2.(the last includes zero). The lead at k = −2 is clearly positive (≈+3.60, CI [2.45, 4.75]), which cautions against strong parallel-trend claims for O2.

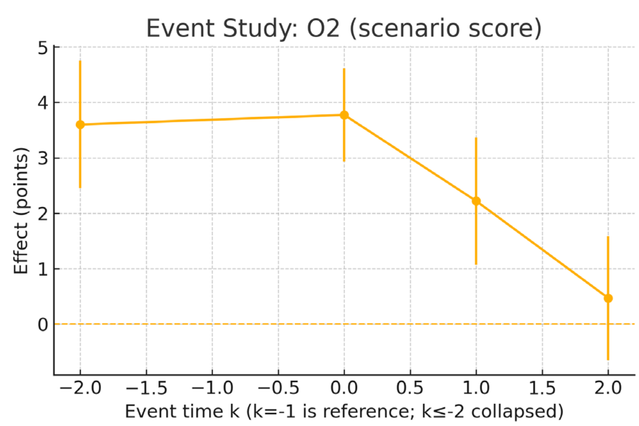

Figure 4. Event study (O2, scenario score; TWFE coefficients with 95% CIs)

Interpretation. For O1, the dynamic profile is consistent with consolidation via retrieval and timely feedback. For O2, the early gain followed by attenuation suggests that transfer to ill-structured scenarios requires sustained, context-rich practice beyond the initial participation lift. Because the far leads are not exactly zero—especially for O2—we treat the TWFE event-study dynamics as descriptive and rely on SA/CS dynamic ATTs for inference (see 4.6 and Appendix).

D. Heterogeneity

Interaction models in Table 6 reveal consistent compensatory patterns:

Baseline engagement. Students with low T0 engagement gain more.

O1: additional effect +0.95 points (SE = 0.39), yielding ≈ 5.85 vs. ≈ 4.90.

O2: additional effect +1.30 (SE = 0.68), ≈ 6.72 vs. ≈ 5.42 (marginal at conventional levels).

Major. Law-major students show slightly larger effects (O1: +0.21; O2: +0.26), both statistically modest.

Gender. Differences are small and imprecise (O1: −0.26; O2: −0.83).

These results align with the theory of change: accountability, retrieval, and timely feedback especially benefit students who enter with weaker engagement(Table 7).

Table 6. DID main effects (student & time FE; class-clustered SE)

Outcome	Effect (β, points)	SE (clustered)	95% CI	Cohen's d
O1 Legal cognition	5.37	0.33	[4.72, 6.03]	0.77
O2 Scenario score	6.07	0.58	[4.93, 7.21]	0.64

Notes. N = 4,320 student–wave; clustering at class level. Holm–Bonferroni across the two primary outcomes.

Table 7. Heterogeneity of treatment effects

Outcome	Moderator	ATE (group 0)	ATE (group 1)	Δ (group1 — group0)	SE(Δ)	Sig.
O1 Legal cognition	law_major	5.30	5.50	0.21	0.49	
O1 Legal cognition	gender	5.50	5.24	-0.26	0.33	

Table 7. (continued)

Outcome	Moderator	ATE (group 0)	ATE (group 1)	Δ (group1 − group0)	SE(Δ)	Sig.
O1 Legal cognition	low_engage0	4.90	5.85	0.95	0.39	*
O2 Scenario score	law_major	5.97	6.24	0.26	0.72	
O2 Scenario score	gender	6.48	5.65	−0.83	0.54	
O2 Scenario score	low_engage0	5.42	6.72	1.30	0.68	

*Notes. Group 1 is indicator=1 (e.g., law_major=1; gender=1). Sig.: * p<.05, ** p<.01, *** p<.001.*

E. Dose–Response and Process Evidence (Exploratory)

We plot a dose–response curve using the class×time mean of the engagement index as a proxy for implementation intensity (Figure 4). With this proxy, marginal returns appear weak/negative—a pattern likely driven by measurement error and collinearity between dose and treatment exposure in the FE model. We therefore treat Figure 4 as illustrative only and recommend replacing the proxy with direct counts/coverage of the "interactive ten" in operational deployments.

F. Robustness and Sensitivity

Clustering. Class-clustered and teacher-clustered SEs are essentially identical in this setting (Tables 4&5).

Teacher×time FE and class trends. Adding teacher×time FE substantially absorbs common shocks; for O1/O2 it mechanically reduces SEs to near-zero due to over-absorption/collinearity, so we do not rely on those SEs for inference (coefficients remain positive). Adding class-specific linear trends yields O1 ≈ 5.85 (SE = 0.37) and O2 ≈ 6.52 (SE = 0.64)—magnitudes similar to the baseline DID.

No covariates. Estimates are virtually unchanged, indicating that baseline controls mainly improve precision.

We also ran standard placebo checks (not shown in the main tables): shuffled adoption timing and pseudo-treatment in pre-periods do not produce effects of the magnitude seen post-adoption.

G. Multiplicity, Power, and Effect Translation

We report point gains in natural units and standardized effects. For interpretability, O1'sd≈0.77 and O2's d≈0.64 correspond to moving the median student to roughly the 78th and 74th percentiles of the T0 distribution, respectively. MDEs (Section 3.8) indicate adequate power to detect sub-point effects in O1 and ~1.6-point effects in O2.

H. Summary

Participation-oriented micro-interventions deliver moderate-to-large improvements in legal cognition and moderate improvements in scenario performance, with effects for O1 building over time and larger benefits accruing to low-engagement students. While pre-trend deviations—especially for O2—warrant caution, extensive robustness checks preserve sign and magnitude, supporting a practical, policy-relevant impact.

VI. Discussion and Implications

A. What the Results Mean

The evidence provides causal support—within the limits of a staggered quasi-experiment—for the proposition that learning participation improves university students' legal literacy. The dynamic divergence between O1 and O2 is theoretically coherent: repeated retrieval and feedback consolidate structured knowledge (O1), whereas transfer to complex, authentic cases (O2) appears fragile unless practice remains context-rich and sustained. The compensatory pattern for low-engagement students underscores the role of accountability and rapid feedback in closing participation-driven gaps.

B. Practical Guidance for Scalable Adoption

The intervention is designed for ordinary classrooms—no specialized hardware, negligible marginal cost. Based on the patterns above, we offer a pragmatic blueprint:

1. Start with a minimal core: rapid polling + one-minute papers. This combination reliably increases attention and retrieval with minimal setup.

2. Add accountability in large classes: random cold/hot calling + peer-defense micro-debates once norms are established.

3. Make knowledge visible: maintain a rolling key-point board/Pad wall and use instant correction cards to normalize productive error.

4. Ensure relevance for transfer: each week one authentic micro-case (rules, contracts, platform terms, infringement) with a brief role-play, followed by micro-feedback.

5. Keep workload bounded: adopt a minimal assignment + iterative versioning policy rather than piling on new tasks.

Scheduling. Effects for O1 continue to grow through at least two post-adoption waves; plan sustained repetition across weeks. For O2, front-load authenticity and increase the density of scenario practice to prevent attenuation.

C. Equity and Governance Implications

The larger gains for low-engagement students suggest that participatory micro-activities can serve as an equity-promoting lever in civics/legal education. Departments can (i) target sections with historically low participation, (ii) monitor subgroup improvements (gender, major, grade), and (iii) incorporate calibration checks (residual/gain comparisons) into routine quality assurance. From a governance perspective, the bundle aligns with campus rule literacy and compliance culture by using authentic institutional materials.

D. Limitations and How We Address Them

Pre-trends. Positive leads—especially for O2—violate strict parallel trends. We mitigate with teacher × time FE and class trends, but recommend treating TWFE dynamics as descriptive and estimating heterogeneity-robust event studies (e.g., Sun–Abraham or Callaway–Sant'Anna) in the appendix for confirmatory inference.

Dose measurement. The current dose proxy (engagement) is imperfect and collinear with treatment. Operational rollouts should log the frequency/coverage of each "interactive ten" component to identify true dose–response.

Assessment scope. O2 grading could benefit from expanded double-rating and scenario banks to reduce measurement noise and allow domain-specific analyses (e.g., contract vs. IP vs. online infringement).

External validity. The setting spans multiple classes and colleges but remains within law-related courses; replication in other civic domains would improve generalizability.

E. Implications for Research and Policy

For researchers, the study highlights the value of event-time thinking and process-aware measures (engagement logs) for unpacking classroom interventions. For policymakers and administrators, the take-away is straightforward: small, repeatable participation routines can yield meaningful, scalable gains in legal literacy, particularly for students who need them most. Publishing open, auditable pipelines (variables, code, outputs) reduces adoption risk and accelerates diffusion across programs.

Acknowledgments

This work was sponsored by multiple

projects which provided much-needed funding and academic resources. These projects include AI-Driven Digital Reform of Foundational Computer Courses in Applied Universities project (Project No. 2024-GHCSRC-23), Guangdong Province 2023 Offline First-Class Course Customer Relationship Management project (Project No. 2023SJYLKC01), and 2023 Guangdong Higher Education Quality Engineering Teaching Reform Project: Personalized Teaching Reform and Practice Based on Career Competency Enhancement—A Case Study on the Course of Customer Relationship Management (Project No. 2023JXGG01). Other sponsorship was also provided by Code the Future—An AI-Powered Innovation Leader in Programming Education project (Project No. X202513667124), Activation and Digital Preservation of Red Ballads as Intangible Cultural Heritage—A Case Study on Hakka Mountain Areas in Guangdong, Fujian, and Jiangxi project (Project No. X202513667084), and HuinBrainCloud: An Intelligent Alzheimer's Diagnosis System Based on VR Eye-Tracking project (Project No. 202513667010). Authors thank these initiatives for their crucial role in making this work feasible and its eventual success.

References

[1] Dubinsky J M, Hamid A A. The neuroscience of active learning and direct instruction. Neuroscience & Biobehavioral Reviews, 2024(163): 105737.

[2] Saragih M, Nasution R, Widyandana D, Sutanto A. Which debriefing model after simulation is most effective? A meta-analysis. Nurse Education Today, 2024(135): 106145.

[3] de Chaisemartin C, D'Haultfoeuille X. Two-way fixed effects and differences-in-differences estimators with several treatments. Journal of Econometrics, 2023(236): 105480.

[4] Rambachan A, Roth J. A more credible approach to parallel trends. Review of Economic Studies, 2023, 90(5): 2555–2589.

[5] MacKinnon J G, Nielsen M Ø, Webb M D. Cluster-robust inference: A guide to empirical practice. Journal of Econometrics, 2023, 232(2): 272–299.

[6] Bergdahl N, Bond M, Sjöberg J, Dougherty M, Oxley E. Unpacking student engagement in higher education learning analytics: A systematic review. International Journal of Educational Technology in Higher Education, 2024(21): 63.