Available Online: 18 November 2026 DOI: 10.54254/3049-7248/2026.39240

When "brain outsourcing" becomes habit: governmentality dilemmas and learning alienation in agentic workflows

Mo Chen*, Min Liao, Yangfeng Zhang, Tong Wang

School of Electronic Information Engineering, Geely University of China, Chengdu, China

*Corresponding Author. Email: 993287003@qq.com

Abstract. The spread of generative AI has made "brain outsourcing"—delegating cognitive tasks to AI—a common phenomenon in higher education, especially in low-tier private undergraduate colleges. Existing studies explain it mainly at the individual level, neglecting the institutional consequences of the technologies used to govern it. Drawing on design-based research principles, this study conducted a three-group quasi-experiment in a Data Structures course at such a college: an adversarial workflow (Group A, n = 36), a negotiative workflow (Group B, n = 40), and a traditional control group (n = 44), triangulated across five chains of evidence. Results show that Group A's direct-dumping rate fell to 11.1% (vs. 57.5% and 75.0%), yet concealed forms—cosmetic modification and division-of-labor outsourcing—rose compensatorily, leaving the composite outsourcing index unchanged across groups (p = 0.864). Process engagement rose sharply while output authenticity stayed low (comment–code matching 3.62/5; only 27.8% of difficulty tags authentic), indicating "performative engagement". Only 1 of 120 students adopted a self-regulated strategy. In Group A, perceived control rose (+2.17) and negotiation space fell (−0.91) without meaningful growth in autonomy; the 18.3-point advantage over the control group in comprehension depth shrank to 6.8 points in transfer application. The "non-outsourceable" workflow thus changed outsourcing's form, not its deep structure, alienating learning into an adversarial game of evading detection—a new hidden curriculum. Agent design should shift from adversarial surveillance to dialogic support.

Keywords: brain outsourcing, agentic workflow, hidden curriculum, performative engagement, private undergraduate college

1. Introduction

1.1. From "AI empowerment" to "brain outsourcing": a concealed institutional paradox

Generative AI is entering higher education in the posture of an "educational transformer", with intelligent tutors, adaptive platforms, and code-assistance tools promising instant feedback, personalized pathways, and vast knowledge supplies—a narrative whose genealogy has been systematically mapped [1]. Yet in the classrooms of a low-tier private undergraduate college, a sharper reality has surfaced: students systematically hand over cognitive tasks—watching videos, solving problems, writing code—to AI, leaving "videos running in the background, problems thrown directly at AI, and zero cognitive investment". This has become a

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

collective norm rather than the opportunism of a few. The term "low-tier private undergraduate college" denotes private institutions at the bottom of admission tiers and social rankings whose students enter with relatively weak academic foundations; it describes the structural position of the research site and carries no value judgment on students.

The evidence on this problem has been accumulating in layers. The first layer answers "does the harm exist?": a systematic review by Zhai et al. shows that over-reliance on AI erodes students' abilities to retrieve information independently and to evaluate it critically [2]. The second layer answers "is the harm equally distributed?": Prather et al. found a "widening gap" effect of generative AI on novice programmers—the weaker one's prior knowledge, the more likely one is to develop the illusory sense of "thinking one has understood", so that the strong get stronger and the weak get weaker [3].

The third layer identifies what makes the problem intractable: "using AI well" itself requires metacognitive monitoring [4]—precisely the capacity that outsourcing erodes. Together, the three layers sketch a vicious circle: the more vulnerable the student, the heavier the reliance on AI, and the greater the loss of the very capacity needed to manage it.

These studies, however, share one analytic frame: they locate the problem in the individual student—who is prone to dependence, what dependence costs, and what capacities individuals should rebuild. What they leave unanswered is this: how does the institutional environment shape the form of AI dependence? And how do the technical means deployed to govern outsourcing, in turn, reshape learning itself?

This study shifts the perspective from the individual to a structural analysis of the "population–institution" nexus. Drawing on Bourdieu's "habitus", Foucault's "governmentality", de Certeau's "tactics of the weak", and the Jackson–Gallagher line of work on the "hidden curriculum", this study seeks to expose a core paradox: a structural rupture exists between the system's design goal of "promoting learning" and students' practical strategies of "evading learning". The more precisely a workflow is engineered, the more precise the evasion tactics students develop—the "effectiveness" of technical governance may constitute its most profound "failure".

1.2. Research question and hypotheses

This study therefore poses one question and three competing hypotheses.

The core question is: when "brain outsourcing" has become the cognitive habitus of students at a low-tier private undergraduate college, does a "non-outsourceable" agentic workflow break that habitus, or does it instead construct structural conditions that force students to "perform learning"? Is such an intervention itself a new hidden curriculum?

H1 (technical-correction hypothesis): A "non-outsourceable" design effectively reduces outsourcing and improves cognitive engagement and learning outcomes.

H2 (habitus-resistance hypothesis): Students will treat "evading the system" as a new goal; outsourcing behavior changes shape but its total volume remains unchanged.

H3 (hidden-curriculum-paradox hypothesis): A "non-outsourceable" design does reduce crude outsourcing, but at the cost of turning learning into an "adversarial game"—students learn "how to appear to be learning" rather than "how to learn".

The three hypotheses are equally falsifiable; most existing research defaults to H1's technical-correction stance, leaving H2 and H3 untested. The analytic weight is placed on H3, which asks whether the very "effectiveness" of technical governance constitutes a new educational problem. H3 is decomposed into five testable sub-hypotheses—H3.1 (outsourcing changes form rather than disappears), H3.2 (engagement becomes performative), H3.3 (evasion becomes collectivized), H3.4 (perceived control rises while autonomy stagnates),

and H3.5 (outcomes meet standards on the surface but fail to transfer)—each corresponding to a chain of evidence tested in Section 4.

2. Literature review and theoretical framework

This chapter assembles the theoretical resources needed to explain why ever-finer technical governance fails to cure brain outsourcing: Section 2.1 takes stock of what existing research has and has not explained; Sections 2.2–2.4 introduce the required lenses in sequence, each entering because the previous one could not finish the job; Section 2.5 links them into the mechanism chain framing the data analysis.

2.1. "Brain outsourcing": what existing research explains, and what it misses

A dominant discourse in educational technology treats generative AI as a "cognitive enhancement tool" that empowers learners through instant feedback and adaptive pathways—a discourse whose genealogy and rhetoric have been systematically documented [1]. Against it, Zhai poses a sharper question: do AI tutors empower learners, or turn them into "cognitive slaves" [5]?

The first line of evidence comes from cognitive psychology: "brain outsourcing" is the contemporary extension of cognitive offloading—transferring tasks such as memory and computation to external media. Risko and Gilbert provide the crucial criterion: offloading becomes alarming when it drifts from "strategic use" to "habitual use", from "assisting thought" to "replacing thought" [6].

A second line answers whether offloading has crossed that line in the age of AI—yes. Zhai et al. establish what happens once the line is crossed: long-term reliance on AI significantly weakens independent retrieval and critical evaluation [2]. Zhang et al. then identify who is more likely to cross it: students with low academic self-efficacy and high academic stress [7]. Together they reveal an inequality: those who depend on AI most can least afford it.

This pattern appears with particular sharpness in computing education, where three studies by Prather's team form a tightening chain: at the outcome level, the "widening gap" effect shows that novices with weak prior knowledge are more prone to "illusory understanding" with AI assistance [3]; at the mechanism level, beginners treat AI as an "omniscient authority", lacking the metacognitive capacity to question and debug it [8]; and at the literature level, the team warns that research focuses too much on technical performance and too little on how AI reshapes cognitive-development trajectories [9].

Together, the three lines explain what outsourcing is, what harm it does, and who suffers most; but all locate the problem in the individual and prescribe individual capacity-building (e.g., the metacognitive monitoring emphasized by Tankelevitch et al. [4]). Three questions remain open: how does a technical governance system operate on students, how do students respond, and why is the response pattern so stable? A clue comes from Reicherts et al.: when AI is perceived as a "surveillance tool" rather than a "dialogue partner", users redirect cognitive resources from "understanding the task" to "understanding the system's rules" [10]—the answer lies in the interaction between the governance system and its subjects, which the next three sections take up.

2.2. Governance and counter-governance: how the system governs, how students evade, and why the pattern persists

2.2.1. Foucault's "governmentality": how the system leads students to police themselves

For Foucault, modern power works most effectively not through naked coercion but through "subtle techniques" that induce the governed to produce obedient behavior on their own; Beattie shows that

performance indicators and behavioral monitoring in higher education produce "docile bodies" [11], lending direct legitimacy to analyzing agentic workflows through governmentality.

The agentic workflow in this study is a digital realization of this logic: "process visualization", "social visibility", "just-in-time generation", and "random teacher spot-checks" form a network of micro-power within which students learn to "self-inspect"—monitoring dwell time, revising comments to match code, and tagging "difficulties" to avoid being returned. The subtlety of this self-discipline is that students often do not perceive themselves as obeying; they sincerely believe that "I am studying hard".

The governmentality lens shows how power operates, but it writes students as too compliant—if governmentality were really as effective as described, outsourcing should have disappeared rather than mutated. To explain students' reactions, a second lens is needed.

2.2.2. De Certeau's "tactics of the weak": how students counter the system

Unlike Foucault, de Certeau turns to the dominated: how the weak carry out everyday resistance through "tactics" within spaces delimited by the strong—poaching in the cracks, surviving on timeliness and concealment. Two studies by Rovea provide a stepped basis for bringing this theory into classrooms: students' everyday learning practices are indeed "appropriations" of the official curriculum [12], and the tactics of disadvantaged groups are highly context-adaptive and collectively shared [13]—organized and transmissible rather than scattered improvisation. The latter point supplies the theoretical expectation for the collective evasion forms (e.g., the "dormitory division of labor") presented in Section 4.

Applied to "brain outsourcing", "direct dumping", "idling", and "copy-pasting" should not be hastily moralized as "laziness" but understood as tactics of the weak—a survival wisdom under academic and assessment pressures. Tactics also evolve: once the system blocks "direct dumping", more concealed and organized forms emerge, such as "cosmetic modification" and "division of labor" (Section 4.1).

The tactics lens explains how students evade, but not why the tactics are strikingly similar across dormitories and cohorts—where they come from and why they persist requires a third lens.

2.2.3. Bourdieu's "habitus": why evasion patterns are passed down

Habitus is not an individual's free choice but "a set of durable dispositions formed by long-term occupation of a particular social position"—a "structured structure" that reproduces the structure that shaped it. Three studies support this application: Mu and Dooley demonstrate the concept's applicability in Chinese educational contexts [14]; Oscar E and colleagues confirm from the student side that class-based habitus shapes choices of university and field of study [15]; and Heather M shows from the teacher side how teacher habitus reproduces social inequality through everyday pedagogic work [16]. Habitus operates at both ends of the classroom.

Viewed through this theory, "brain outsourcing" is not a matter of individual morality but a collective cognitive habitus. Students at low-tier private colleges have repeatedly experienced "working hard yet achieving little", internalizing the dispositions of "I can't learn" and "learning is useless"; when generative AI appeared with "low thresholds and high returns", they did not "choose" outsourcing but activated the minimal-effort strategy almost unconsciously. Any technical design that dares not confront this deep structure will ultimately be absorbed, distorted, or circumvented by it.

The three lenses now cover power (governmentality), response (tactics), and stability (habitus); what remains is the educational consequence—in this game of governance and counter-governance, what do students ultimately "learn"?

2.3. The hidden curriculum: what does this game teach students

Jackson's classic account holds that what students "really learn" in school is often tacit knowledge about rules, power relations, and survival strategies. Two strands of research have consolidated the concept: a review by

Meyer et al. shows that the hidden curriculum is pervasive and dual in function, serving professional socialization while also transmitting harmful norms [17]; and Nina R and Liezel F show that in marketized environments it systematically obscures knowledge unfavorable to capital accumulation, becoming an ideological tool that maintains power structures [18]—the latter concerning precisely private higher education, the closest match to the present research site.

These studies, however, analyze institutions of "persons and persons"; applying the hidden curriculum to workflows of "persons and algorithms" requires a bridge, which Gallagher et al. provide with the concept of "ontological transparency": educational technology transmits not only disciplinary content but, through interface design, interaction logic, and evaluation mechanisms, tacitly defines "what learning is" [19].

Applying this insight to agentic workflows reveals that the "non-outsourceable" design is producing a new hidden curriculum. When a workflow requires students to "expose traces of thinking", "tag difficulty points", and "write causal comments", it actually teaches that "learning" is not exploring the unknown but submitting acceptable evidence to a surveillance system, and that "understanding" is not an inner mental state but an algorithmically detectable behavioral trace. Students immersed in this curriculum gradually equate "learning" with "successfully evading detection" and "thinking" with "appearing to think"—the theoretical source of this study's "hidden curriculum paradox": the harder governance works, the more advanced the "evasion course" students master.

Could "better technology" bypass this paradox? Two strands of literature say no: Margaret B and colleagues diagnose a "techno-solutionist" myth in AI-in-education discussions [1], locating the problem in the logic of application rather than the technology itself; and normative research identifies the alternative—Naya J, rereading Freire, insists that genuine education is "dialogic" [20], and Peuler and Harlow stress that critical instructional design must democratize power relations [21], both foreshadowing the "dialogue over confrontation" recommendation in Section 6. In the view of this paper, the adversarial "non-outsourceable" design is essentially a digitalized banking education: the system stores the rules and judges compliance while students passively adapt—not co-constructors of knowledge but decoders of rules and evaders of detection.

2.4. Desirable difficulties: how the cognitive cost is incurred

The four lenses above explain power, behavior, and educational consequences; one cognitive-level question remains unanswered. The workflow was designed to create "desirable difficulties" that would promote deep learning—so why was difficulty created while learning did not occur?

The evidential basis of desirable-difficulty theory has been consolidated layer by layer: the Bjorks' laboratory studies show that learning activities promoting deep understanding must involve moderate cognitive friction, such as the generation, spacing, and interleaving effects [22]; H. B. A. B. D. and colleagues translated these findings into the operable S2D2 framework for teachers [23]; and Rutherford verified the contribution to durable learning in the different discipline of design education [24]. The necessity itself is beyond doubt.

While endorsing the educational value of desirable difficulties, this study poses a critical question: if cognitive friction is compulsorily imposed by an external algorithmic system rather than autonomously regulated by the learner, will its effects be alienated? Reicherts et al. have planted the seed: when cognitive support is perceived as a "surveillance tool", users redirect cognitive resources from "understanding the task" to "understanding the system's rules" [10]. The effect of difficulty thus depends on how the learner views the agent imposing it—a question the theory has not addressed and a variable agentic workflow cannot evade.

Extending this finding to educational settings, this study proposes the concept of the "alienation of desirable difficulty". In Group A's workflow, the system manufactures high-intensity cognitive friction through

forced loops, semantic detection, and comment matching; once students redirect cognitive resources from "understanding the principle of linked-list insertion" to "how to write comments that the system will pass", the difficulty is alienated—no longer a "scaffold" for deep understanding but a "cage" of performed suffering. Students learn not "how to think" but "how to display the suffering of thinking", developing not "cognitive capacity" but "suffering-management capacity". When "desirable" becomes "compulsory" and "friction" becomes "performance", education slides toward alienation.

2.5. An integrated theoretical framework: a mechanism chain

The five lenses are strung together not as a parallel list but as an interlocking mechanism chain, each link answering the question left by the previous one (Table 1).

Table 1. The integrated theoretical framework

Level	Theoretical resource	Question answered	Mapping in this study
Technical governance	Foucault	How does the system induce self-discipline?	Algorithmic surveillance, process visualization, random spot-checks
Tactics of the weak	de Certeau	How do students counter the system?	Evasion strategies such as "cosmetic modification", "division of labor", and "segmented copying"
Reproduction of habitus	Bourdieu	Why are evasion patterns stably transmitted?	"Brain outsourcing" solidifies from an individual strategy into a group norm
Hidden curriculum	Jackson–Gallagher	What does this game teach students?	An alienated ontology of "learning = successfully evading system detection"
Alienation of desirable difficulty	Bjork–this study	What is the cognitive cost?	Students learn to "perform suffering" rather than "construct understanding"

The full statement of the mechanism chain is as follows. The workflow exercises control in the manner of governmentality; students enter the system carrying a habitus formed by prior educational selection; habitus drives them to develop evasion tactics; the long-term interaction of governance and evasion transmits to students the hidden curriculum that "learning equals passing detection"; and finally, the desirable difficulty that should have promoted understanding is alienated into an object of performance, leaving cognitive engagement unrealized.

The core proposition of this framework is that the effectiveness of technical governance in "controlling behavior" constitutes precisely its failure in "cultivating learners". This escalating dynamic is not the failure of any single design but the expression of a paradox internal to technical governance: it attempts to cultivate autonomy through control and to emancipate the subject through discipline, and in the end alienates learning into an adversarial game of "successfully evading the system".

This proposition is further supported by two strands of literature. *How People Learn* emphasizes that deep understanding must include the conditionalization of knowledge—knowing when, where, and why to use it [25]; this provides the criterion by which Section 4 tests learning quality through "transfer application"—students who solve problems "within the system rules" but cannot transfer have acquired only over-contextualized surface skills. From a critical standpoint, Marko T cautions that educational technology must continually ask "whose interests are served and whose voices are suppressed" [26]; this framework gives voice to the most silent learners under technical governance—the "brain outsourcing" population of low-tier private undergraduate colleges.

3. Research design

3.1. Research site: why a low-tier private undergraduate college

The study was conducted at a low-tier private undergraduate college in China—an "extreme case" within purposive sampling: a "non-outsourceable" workflow should be tested where the outsourcing habitus is most severe. Lower-ranked private colleges concentrate students systematically marginalized by prior educational selection [27], among whom low academic self-efficacy and scarce success experiences have turned "throwing problems directly at AI, idling videos, and zero cognitive investment" from an individual strategy into a group norm.

3.2. Participants and grouping: a three-group quasi-experiment

The study drew on Design-Based Research (DBR) principles, which suit testing instructional interventions in real classrooms [28]; as only one "design–implementation–analysis" cycle was completed within a single semester, it is more accurately a classroom quasi-experiment informed by DBR. Conducted in the first semester of the 2025–2026 academic year in a Data Structures course, it comprised three groups: Group A (adversarial workflow—"blocking" direct AI answers), 36 students from Class 1 of the 2024 Electronic Information cohort; Group B (negotiative workflow—preserving outsourcing while raising its cost), 40 students from the 2024 Artificial Intelligence cohort; and a control group (traditional lecturing, homework, and final exam, with no agent intervention), 44 students from Class 2 of the 2024 Electronic Information cohort. Artificial Intelligence majors, more familiar with AI tools, test whether gentle channeling holds against "tech veterans" (Group B); the weaker-programming Electronic Information classes maximize the test of "blocking" intensity given the "widening gap" effect [3] (Group A). Disciplinary background and treatment are fully confounded in this design (Section 6.4).

3.3. The intervention: the design logic of two workflows

Both workflows start from desirable-difficulty theory but take opposite routes: Group A imposes difficulty algorithmically, Group B through guided negotiation. Group A followed five "anti-outsourcing" principles: (1) process visualization and path locking (sequential node unlocking, undraggable videos, full logging); (2) contextual anchoring (problems embedded in randomly generated in-class code); (3) just-in-time generation (randomized parameters preventing proxy completion); (4) social visibility (peer-visible outputs as a behavioral constraint [29]); and (5) cognitive friction (reasoning nodes that cannot be completed with one click).

Each Group A session comprised five progressive nodes: (1) contextualized diagnosis—students explain in voice within five minutes where a randomly generated buggy code segment fails, with semantic analysis detecting causal reasoning rather than keywords; (2) generative concept mapping—students draw a relation map of pointer concepts and mark the relationship they understand least; (3) adversarial programming—students complete randomized incomplete code with three lines of causal-reasoning comments; (4) peer-visible reflection—each student comments on at least one peer's anonymized output; and (5) random teacher spot-check—20% of students are randomly selected for a brief real-time dialogue, creating an unpredictable deterrent.

Group B shared the same random code generation and task structure but left room for negotiation at every node: AI hints viewable with mandatory annotation ("what the AI suggested and whether I agree"); a modifiable concept-map template; thinking hints and a reference answer followed by a written "why I didn't

think of it"; optional peer reflection with bonus credit; and an agent-generated "learning diagnosis report" replacing the teacher spot-check. The intent was not to prohibit outsourcing but to raise its metacognitive cost.

3.4. Data collection and analysis: five chains of evidence and statistical methods

An explanatory sequential mixed-methods design was adopted: quantitative analysis identified between-group differences, and qualitative analysis explained why and how they arose.

3.4.1. Five-chain evidence matrix

Because no single data source can distinguish genuine learning from performed learning, a five-chain triangulation matrix was built (Table 2), letting multiple data sources interrogate one another.

Table 2. The triangulation matrix of the five chains of evidence

Evidence chain	Data sources	Core function	Key indicators
A. Outsourcing behavior	Platform tracking logs + assignment detection	Objective behavioral traces	Direct-dumping rate (A1), video completion rate (A2), cosmetic-modification rate (A3), division-of-labor rate (A4)
B. Cognitive engagement and output authenticity	Process logs + algorithmic detection	Distinguishing genuine thinking from feigned attempts	Node dwell time (B1), revision attempts (B2), comment–code matching (B3), authenticity of difficulty tags (B4)
C. Strategy types	Platform-log coding + semi-structured interviews and focus groups (subsample cross-validation)	Revealing the micro-mechanisms of evasion tactics	Type distribution from open → axial → selective coding; number of strategy switches
D. Power perception	Pre/post scales	Changes in perceived control, negotiation, and autonomy	Perceived control (D1), negotiation space (D2), autonomy (D3)
E. Learning outcomes	Contextualized pre/post tests	Testing in-system attainment vs. out-of-system transfer	Comprehension depth (E1), transfer application (E2), metacognitive calibration (E3)

3.4.2. Measurement specifications

All three groups used the same online platform; the control group, though without workflow nodes, received identical video tasks and homework, making A2 comparable across groups. A3 and A4 were determined jointly from homework textual similarity, AST-based structural comparison, and interviews. B1 and B3 are workflow-specific, compared only between Groups A and B. A1 is a student-level binary indicator (any direct dumping during the semester), distinct from the Chain C share of "direct dumpers" classified by dominant strategy. A composite outsourcing index (sum of the standardized behavioral indicators) was constructed to test differences in total volume.

3.4.3. Reliability and baseline

Chain D subscale reliabilities (pre-test/post-test): perceived technical control (D1, 5 items), α = 0.760/0.930; negotiation possibility (D2, 5 items), 0.737/0.826; learning autonomy (SRQ-A, D3, 15 items), 0.880/0.897;

metacognitive monitoring (12 items), 0.852/0.843. Baselines were equivalent across groups (course pre-test $F(2,117) \approx 0.00$, $p > 0.05$; SES pre-test $F(2,117) = 1.62$, $p = 0.203$), and the whole-sample SES mean of 2.23 (5-point scale) corroborates antecedent habitus. Chain E uses post-test scores as dependent variables.

3.4.4. Statistical methods

Categorical between-group comparisons used the chi-square test, with Cramér's V as the effect size (Equations (1) and (2)):

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{1}$$

$$V = \sqrt{\frac{\chi^2}{N(k-1)}} \tag{2}$$

where O and E are the observed and expected frequencies and k is the smaller of the row and column counts; sparse cells (expected frequency < 5) were merged for robustness retesting. Three-group continuous comparisons used one-way ANOVA, with η^2 as the effect size (Equations (3) and (4)):

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} \tag{3}$$

$$\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}} \tag{4}$$

with Tukey's HSD for post hoc comparisons. Two-group continuous comparisons (Groups A and B) used independent-samples t tests, with Cohen's d as the effect size (Equation (5)):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{1/n_1 + 1/n_2}}, \qquad d = \frac{\bar{X}_1 - \bar{X}_2}{s_p} \tag{5}$$

where s_p is the pooled standard deviation. Within-group pre–post changes used paired-samples t tests (effect size $d_z = \bar{D}/s_D$). For null results such as unchanged autonomy, TOST equivalence testing (bounds ± 0.3) was added: a change was deemed negligible only when both one-sided hypotheses ("change ≤ −0.3" and "change ≥ 0.3") were rejected. Section 4 cites these equations by number.

3.4.5. Qualitative analysis

Two independent coders followed Strauss and Corbin's grounded-theory procedure (open → axial → selective coding); double-coding agreement reached $\kappa = 0.72$ for B4 difficulty-tag authenticity and 90% for Chain C strategy classification (interview subsample of 10 per group versus platform-log coding). Integration followed "mutual interrogation": when logs showed "completion" while interviews revealed "idling plus copying", that rupture itself served as core evidence of performative engagement.

3.5. Research ethics

As a teaching-intervention study in regular classroom instruction involving no additional risks, the study fell outside the host institution's mandatory ethics review. Protective measures were nonetheless implemented: all students signed informed-consent forms covering the research use of platform logs, assignment texts, and interview recordings, with free withdrawal and no impact on grades; all data were anonymized and are reported only in aggregate; and transcription began only after course assessment ended. The study adhered to the principles of the Declaration of Helsinki. Student quotations are transcript excerpts coded as PT (individual interview) / FG (focus group) / DI (teacher spot-check dialogue) – group – serial number.

4. Findings

The five sub-hypotheses of H3 are tested through five chains of evidence, each reported in a common rhythm: question, quantitative results, micro-operations from interviews, verdict.

4.1. Chain one: the morphing of outsourcing behavior (H3.1)

Did Group A's outsourcing really decrease? Its direct-dumping rate was only 11.1%, versus 57.5% in Group B and 75.0% in the control group—a significant difference ($\chi^2(2, N = 120) = 33.68$, $p < 0.001$, Equation (1)) with a large effect size (Cramér's $V = 0.53$, Equation (2)); the video completion rate also rose to 0.73 through path locking (Table 3, upper part). By these indicators alone, the "non-outsourceable" workflow appears to suppress the crudest outsourcing.

Table 3. Descriptive statistics of outsourcing-behavior indicators across the three groups

Indicator	Group A (adversarial)	Group B (negotiative)	Control (traditional)
A1 Direct-dumping rate	11.1%	57.5%	75.0%
A2 Video completion rate	0.73 (0.11)	0.48 (0.17)	0.33 (0.17)
A3 Cosmetic-modification rate	0.70 (0.21)	0.39 (0.19)	0.31 (0.15)
A4 Division-of-labor rate	0.58 (0.17)	0.39 (0.19)	0.29 (0.18)
A5 Composite outsourcing index	1.81 (0.50)	1.82 (0.60)	1.87 (0.52)

Note: A1 is the share of students who ever engaged in direct dumping during the semester (Section 3.4); A2–A5 are means (SD); the control group's A2 uses after-class online video tasks, and A3–A4 were determined jointly from homework detection and interviews.

The lower part of Table 3, however, tells another story. Concealed forms rose compensatorily: Group A's cosmetic-modification ($M = 0.70$) and division-of-labor ($M = 0.58$) rates significantly exceeded the other groups' (Equation (3): $F(2,117) = 49.35$ and 25.49, both $p < 0.001$; $\eta^2 = 0.46$ and 0.30, Equation (4); Tukey's HSD: A > B and control). Meanwhile, the composite outsourcing index was nearly identical across groups (1.81, 1.82, 1.87; $F(2,117) = 0.15$, $p = 0.864$): total volume was conserved. Blocked from "direct dumping", evasion migrated to more concealed, collaborative forms rather than decreasing.

Process-tracing interviews reveal the micro-operations behind this substitution (PT-A-10):

"I had the AI write the code, but I didn't dare submit it directly. I replaced all the variable names with ones I normally use, reordered some statements, and edited the comments on top of the AI's comments. I know the system checks whether the comments match the code".

Juxtaposed with the platform log ("code submitted—comment–code matching 3.62—judged as passed"), this self-report reveals deliberate detection evasion: cognitive resources were invested in the judgment rules themselves.

Verdict: H3.1 is supported—outsourcing morphed rather than disappeared; total volume was conserved.

4.2. Chain two: the performativity of cognitive engagement (H3.2)

Was Group A's invested time genuine thinking or performance? On the engagement side (Table 4), Node 3 dwell time reached 25.4 minutes and revision attempts averaged 3.2, both far above Group B (Equation (5); all $p < 0.001$; $d = 0.93$–1.97). On the authenticity side, however, comment–code matching averaged only 3.62 —"partial match", mentioning code behavior without touching causal mechanisms—and only 27.8% of difficulty tags were highly authentic, the rest being generic statements. High process indicators coupled with

low authenticity define "performative engagement": time was spent "satisfying the system", not "making the brain understand".

Table 4. Descriptive statistics of cognitive-engagement and output-authenticity indicators

Indicator	Group A	Group B	Control	t(74)	Cohen's *d*
B1 Node 1 dwell time (min)	8.7 (2.6)	5.3 (1.9)	—	6.41***	1.47
B1 Node 2 dwell time (min)	9.4 (2.6)	6.7 (2.4)	—	4.79***	1.10
B1 Node 3 dwell time (min)	25.4 (6.0)	19.2 (4.4)	—	5.25***	1.21
B2 Revision attempts	3.2 (1.3)	1.0 (0.9)	—	8.56***	1.97
B3 Comment–code matching (1–5)	3.62 (0.74)	2.97 (0.66)	—	4.04***	0.93
B4 Share of highly authentic difficulty tags	27.8%	22.5%	—	—	—

Note: *** indicates statistical significance at the 0.1% level; independent-samples t tests between Groups A and B (n = 36/40). B3 is the agent's semantic rating (5-point), correlated at r = 0.89 with manual re-rating of the 20% spot-check sample. B4 is the share of coder-rated "high authenticity" (top of 3 levels; double-coding κ = 0.72), not parametrically tested as a proportion. The control group had no workflow nodes.

Interviews show that this performance was self-aware (PT-A-07):

"For that 'what I understand least' tag, I just write something random—like 'pointers are a bit hard to use'—which definitely won't get sent back. Writing something too specific only invites trouble. What if the teacher asks about it? Wouldn't that give me away?"

Students were not merely performing but calculating a safe dosage—tagging a difficulty to pass detection while avoiding tags specific enough to invite follow-up questions: a design meant to expose cognitive obstacles became a calibrated risk-management act.

Verdict: H3.2 is supported—high engagement indicators with low authenticity: "front-stage performance, backstage emptiness".

4.3. Chain three: the collectivization of evasion strategies (H3.3)

Was evasion individual, or had it solidified into a group norm? Strategy-type distributions coded from platform logs (seven categories; reliability in Section 3.4) (Table 5) differed significantly across the three groups ($\chi^2(12, N = 120) = 35.00$, $p < 0.001$, Equation (1)); a robustness test with categories merged into three macro-classes held ($\chi^2(4) = 27.51$, $p < 0.001$). Two numbers suffice: in Group A, cosmetic modifiers (50.0%) and division-of-labor outsourcers (16.7%) summed to 66.7%—two-thirds chose evasive forms requiring collaboration and disguise—while the control group was dominated by direct dumping (50.0%). Group A also averaged 2.97 strategy switches per person, far above Group B (1.45) and the control group (0.70) ($F(2,117) = 33.01$, $p < 0.001$, $\eta^2 = 0.36$, Equations (3)–(4)): strong surveillance accelerated probing and migration among strategies (Figure 1).

Table 5. Distribution of strategy types across the three groups (%)

Strategy type	Group A (n = 36)	Group B (n = 40)	Control (n = 44)
Direct dumping	11.1	27.5	50.0
Idle-cramming	2.8	10.0	20.5
Cosmetic modification	50.0	40.0	9.1
Division of labor	16.7	5.0	9.1

Table 5. Continued

Probing	5.6	10.0	9.1
Struggling	11.1	7.5	2.3
Self-regulated	2.8	0	0

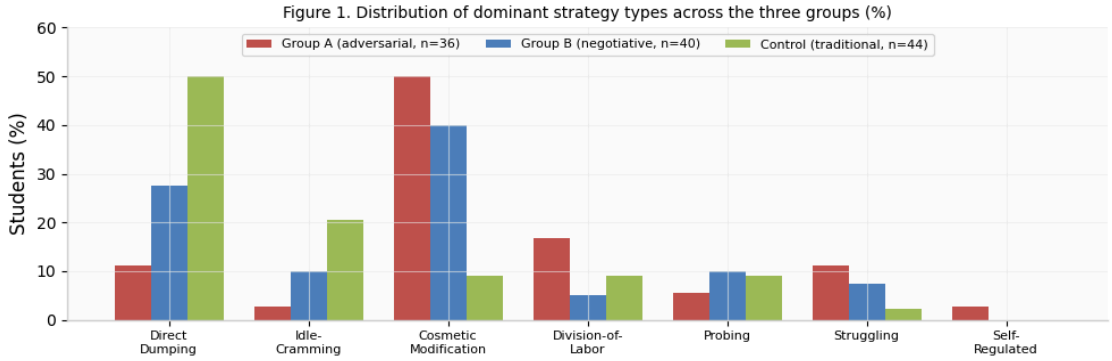


Figure 1. Distribution of dominant strategy types across the three groups (%)

Focus groups revealed the collective organization behind these strategies: multiple groups independently converged on the "dormitory division of labor" (each member outsourcing different nodes), "segmented copying" (interleaving AI output with personal phrasing), and "variable-name substitution" (FG-A-03):

"There are four of us in our dorm. Each week, whoever's turn it is asks the AI about the nodes first and posts the usable parts to our group chat. The rest of us revise them a bit and submit. It was never really discussed; everyone just started doing it".

"Everyone just started doing it": evasion spread as a self-evident group norm without explicit negotiation—habitus manifesting in micro-interaction. Only 1 of the 120 students (0.8%, in Group A) exhibited a self-regulated strategy: neither intervention route cultivated "independent thinking first, AI verification second" at scale within one semester.

Verdict: H3.3 is supported—evasion solidified into group-shared norms, and self-regulated strategies nearly vanished.

4.4. Chain four: the contradictoriness of power perception (H3.4)

Table 6. Pre–post change in power perception (post − pre)

Indicator	Group A	Group B	Control
D1 Perceived control	+2.17 (1.01)***	+0.96 (0.76)***	−0.05 (0.79)
D2 Negotiation space	−0.91 (0.65)***	−0.18 (0.74)	−0.11 (0.74)
D3 Autonomy	−0.17 (0.39)*	+0.10 (0.58)	−0.04 (0.51)

Note: *** and * indicate statistical significance at the 0.1% and 5% levels respectively. Although Group A's D3 change reached statistical significance, TOST equivalence testing placed it within the ± 0.3 equivalence bounds, i.e., of no practical significance; changes in Group B's negotiation space and in all control-group indicators were nonsignificant.

What changed in the subjective experience of students "managed" by the system? Table 6 presents the pre–post change scores. Perceived control soared in Group A ($t(35) = 12.88$, $p < 0.001$, $d_z = 2.15$; between-group

F(2,117) = 66.67, $p < 0.001$, $\eta^2 = 0.53$). Negotiation space collapsed (t(35) = −8.42, $p < 0.001$, $d_z = -1.40$; F(2,117) = 14.50, $p < 0.001$, $\eta^2 = 0.20$). Autonomy stagnated: the change was statistically significant (t(35) = −2.60, $p = 0.014$) but small ($d_z = -0.43$), and TOST equivalence testing (bounds ± 0.3, Section 3.4) placed the 90% confidence interval [−0.28, −0.06] entirely within the equivalence bounds—statistically "changed", practically "unchanged" (between-group F(2,117) = 2.73, $p = 0.069$). Figure 2 shows the three changes concentrating in Group A: a "control paradox" in which feeling controlled did not grow self-management.

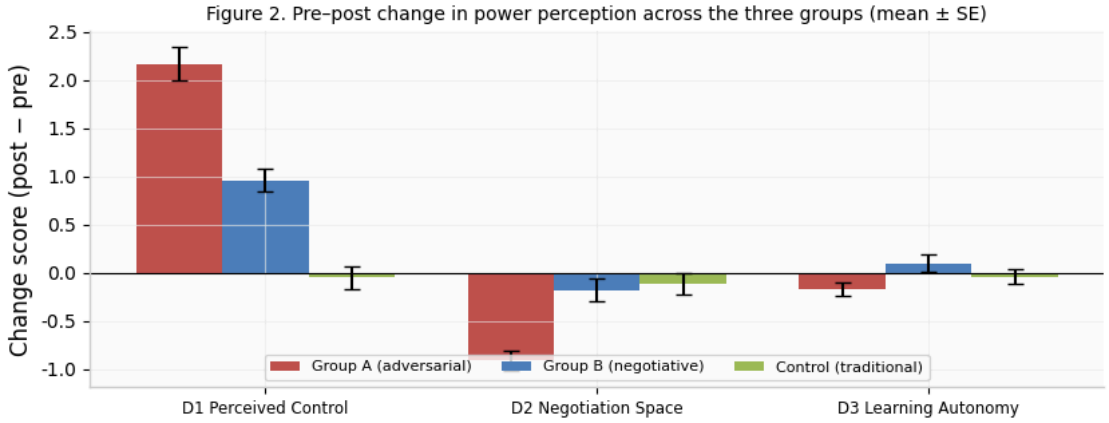


Figure 2. Pre–post change in power perception across the three groups (mean ± SE)

One Group A student conveyed the texture of this control (PT-A-03):

"It feels like it's watching you all the time. If it sends your work back, you have to rewrite it. You never know exactly what it wants, so you just keep trying, again and again, until it's satisfied. There's no room to negotiate".

"Trying until it's satisfied" and "no room to negotiate" footnote the scale data: students felt controlled (D1 rising) and experienced the system as unnegotiable (D2 falling), yet converted none of it into self-management (D3 unchanged)—they learned "surviving under management", not "managing themselves".

Verdict: H3.4 is supported—soaring perceived control, collapsed negotiation space, and stagnant autonomy concentrated in Group A.

4.5. Chain five: the surface attainment of learning outcomes (H3.5)

Was Group A's "better learning" real beyond the system? Table 7 presents post-test outcomes (baselines equivalent, Section 3.4), read in three layers. Comprehension depth: Group A led (62.0 vs. 54.7 and 43.7; F(2,117) = 34.45, $p < 0.001$, $\eta^2 = 0.37$; 18.3 percentage points above the control). Transfer application: the advantage shrank sharply (F(2,117) = 7.43, $p < 0.001$, $\eta^2 = 0.11$); Groups A (34.4) and B (34.7) did not differ ($p = 0.989$), together exceeding the control (27.6) by only about 7 points. Interaction test: computing Δ = E1 − E2 for each participant, a between-group ANOVA on Δ (Equation (3)) confirmed a significant task-type × group interaction (F(2,117) = 7.35, $p = 0.001$, $\eta^2 = 0.11$): the 18.3-point comprehension gap versus the 6.8-point transfer gap indicates contextualized skill "within the system rules" rather than transferable understanding (Figure 3). Notably, Group B matched Group A on transfer at a much lower perceived-control cost (developed in Section 6.3).

Table 7. Descriptive statistics of post-test learning-outcome indicators across the three groups (%)

Indicator	Group A	Group B	Control	Test statistic	Effect size
E1 Comprehension depth	62.0 (10.4)	54.7 (8.2)	43.7 (11.0)	$F(2,117) = 34.45^{***}$	$\eta^2 = 0.37$
E2 Transfer application	34.4 (9.9)	34.7 (10.0)	27.6 (8.8)	$F(2,117) = 7.43^{**}$	$\eta^2 = 0.11$
E3 Share of high calibration	13.9%	10.0%	9.1%	$\chi^2(4) = 0.66$ n.s.	—
E3 Share of low calibration	50.0%	55.0%	56.8%	—	—

Note: *** and ** indicate statistical significance at the 0.1% and 1% levels respectively; n.s. = not significant. Tukey's HSD for E1: A > B > C; for E2: A ≈ B > C (A vs. B, $p = 0.989$). E3 is the share within a three-level rating (low/medium/high calibration); the three groups' distributions did not differ significantly ($p = 0.956$).

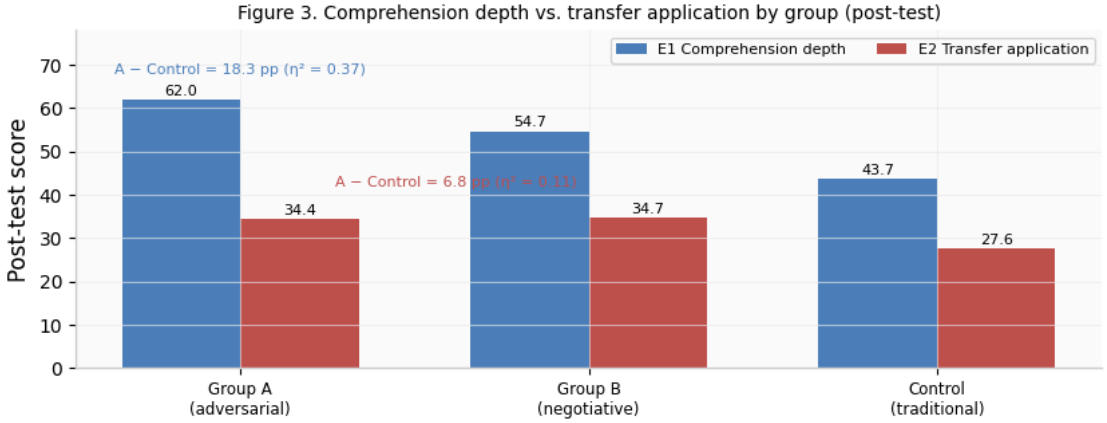


Figure 3. Comprehension depth vs. transfer application by group (post-test)

A Group A student with a high comprehension-depth score, asked to explain the key steps of linked-list insertion without notes (PT-A-04):

"In that system, I can definitely do it—I know the steps by heart. But if you ask me to explain it right now, off the top of my head… I roughly know you find the position first and then change the pointers, but as for which pointer moves at which step, I'd need to look at the code to explain it clearly".

"Only within the system" and "only while looking at the code": knowledge remained bound to context. Metacognitive calibration echoes this—high-calibration shares were low in all groups (13.9%, 10.0%, 9.1%), low-calibration shares exceeded half (50.0%–56.8%), with no significant between-group difference: even Group A students who "performed better" mostly could not judge whether they had truly understood.

Verdict: H3.5 is supported—attainment within the system, mediocrity beyond it, and students were unaware of the difference.

4.6. Integrated test of H3: the five chains converge on the hidden-curriculum paradox

Table 8 integrates the five chains, and the three competing hypotheses receive asymmetric verdicts. H1 (technical correction) is rejected: outsourcing persisted through form substitution (Chain One), engagement authenticity was insufficient (Chain Two), and outcome advantages were confined to in-system contexts (Chain Five). H2 (habitus resistance) receives direct support—the conserved composite index ($p = 0.864$) and evasion solidified into group norms—but cannot explain why resistance took a performative form or why control perception soared while autonomy stagnated (Chains Two and Four). H3 (hidden-curriculum paradox)

is therefore established as the explanatory deepening of H2: resistance became "performance" because the workflow itself set "appearing to learn" as the passing standard—required traces, difficulty tags, and causal comments were met with calibrated traces, safe tags, and matching comments, redefining "learning" as "submitting evidence the system will accept". Effectiveness in controlling behavior constitutes failure in cultivating learners: the workflow changed the surface form of "brain outsourcing" without breaking its deep structure, and what was learned most thoroughly was not data structures but evasion itself.

Table 8. H3 hidden-curriculum paradox: verification across the five chains

Hypothesis	Core data	Verdict
H3.1 Outsourcing morphed	Direct-dumping rate 11.1% (control 75.0%); cosmetic modification + division of labor = 66.7%; composite outsourcing index showed no between-group difference (p = 0.864)	Volume conserved; form substituted
H3.2 Performative engagement	Node 3 dwell time 25.4 min, yet comment matching only 3.62 and high authenticity only 27.8%	Front-stage performance, backstage emptiness
H3.3 Collective evasion	Cosmetic modification + division of labor = 66.7%; strategy switches 2.97, the highest of the three groups; only 1/120 self-regulated	Evasion solidified into a group norm
H3.4 Control paradox	Perceived control +2.17; negotiation space −0.91; autonomy −0.17 (within equivalence bounds)	Disciplined but not autonomous
H3.5 Surface attainment	Comprehension gap 18.3 pp vs. transfer gap only 6.8 pp (37.2% of the former); interaction p = 0.001	Attainment in-system, mediocrity out-of-system

5. Discussion

5.1. The technical translation of governmentality and its boundary

Governmentality in Foucault's sense relies not on prohibition but on environmental design, classification, and technologies of the self that induce "voluntary" compliance; the agentic workflow translates this logic technically through planned node sequences, algorithmic scoring, and panoptic logging. Superficially it "succeeded" (direct dumping 75.0% → 11.1%; Node 3 dwell time 25.4 minutes). But power inevitably provokes counter-conduct: students redirected cognitive resources toward the governance system itself—analyzing detection rules, calibrating performance dosage, organizing collective divisions of labor—while total evasion was conserved (p = 0.864). The workflow is thus not a failed case of governmentality but a case of its "over-success": behavior coupled tightly to system goals at the level of traces rather than cognition. Students learned to produce "the appearance of learning" and the system to recognize it, closing what is termed here the "surface closed loop of governance"—a self-sufficient exchange around behavioral indicators that hollows out the educational goal.

5.2. Tactics of the weak: a de Certeauian moment in the digital age

De Certeau distinguishes the territorial "strategies" of the strong from the "tactics" of the weak—"guerrilla warfare on another's territory". The students' evasion is a digital-age specimen: the "dormitory division of labor" converts individual pressure into a collective network, "segmented copying" exploits the granularity boundary of semantic detection, and "variable-name substitution" turns required "personalized" traces into a

mass-producible procedure—none challenging the system head-on, all turning its rules into resources. Group A's 2.97 switches per person (versus 1.45 in Group B and 0.70 in the control) show that strong control accelerated rather than suppressed tactical iteration. If such creativity were directed at knowledge construction rather than evasion, considerable learning potential would be released—a premise for Section 6.3.

5.3. The reproduction of habitus: Bourdieu's warning

Habitus—the "structured structure" internalized from past experience—explains the pattern's stability: students at this site mostly carry long experience of "learning failure", making "credits at minimal investment" an unreflective tendency, as the baseline SES mean of 2.23 (5-point scale) confirms. A single semester of intervention confronted a habitus shaped by more than a decade of educational selection; that only 1 of 120 students (0.8%) exhibited self-regulation is the quantitative footnote to its stubbornness. Bourdieu's "hysteresis" is also confirmed: the workflow changed field conditions (no direct dumping, mandatory traces), but habitus did not adjust in step—students applied minimal investment to compulsory traces, producing "performative engagement": a misaligned adaptation to the new field's "rules" without entering its "game".

5.4. The hidden-curriculum paradox: technical teaching in Jackson's framework

Extending Jackson's hidden curriculum to agentic workflows reveals that the detection logic itself is a course no one teaches yet everyone must take. Over four months, Group A students arguably learned four hidden lessons better than data structures: detection has boundaries (perform within, relax beyond); difficulty tags require a "safe dosage"; collaboration works if traces are dispersed; and "causal reasoning" can be produced formally (keywords, word count, behavioral match). Absent from any syllabus, this content is mastered because it ties directly to immediate interests—passing, avoiding returns, dodging spot-checks; transmitted through algorithmic judgment, instant feedback, and unpredictable deterrence, it is drilled more frequently and precisely than Jackson's daily rituals. This is the "hidden-curriculum paradox": designed to transmit "deep learning", the workflow actually transmits "evading detection", and every explicit success adds bricks to tacit educational failure.

5.5. The alienation of desirable difficulty: from cognitive psychology to governance technology

Desirable-difficulty theory [22] presupposes that learners genuinely undergo the difficulty; when difficulty is outsourced to AI, divided among roommates, or performed for the system, it is bypassed rather than undergone. The comprehension–transfer gap (18.3 vs. 6.8 points) shows in-system problem-solving that never became transferable structure: the difficulty existed in the task, not in the learner's experience. Encoded into an algorithm and externally imposed, desirable difficulty is alienated from a cognitive-psychological concept into a governance technology—the rational response shifts from "endure for long-term gain" to "evade to cut immediate cost", because immediate system feedback shapes behavior more powerfully than distant gains. External cognitive-load regulation takes effect only with learners' metacognitive participation [10]; this study proves the converse: compulsory difficulty without it is systematically alienated into performance.

5.6. Summary: the stacking of three alienations

Three mutually reinforcing alienations emerge. Behavioral alienation: outsourcing morphed into "cosmetic modification" and "division of labor", with total volume conserved. Cognitive alienation: engagement morphed into "performative engagement", with traces present but understanding absent. Curricular alienation: the learning goal morphed from mastering knowledge into passing detection, the hidden curriculum replacing the explicit one. Once "brain outsourcing" becomes habitus, technical "blocking" meets creative resistance; the

finer the governance, the more exquisite the resistance—a "governance–evasion" race in which the real loser is neither the system nor the students, but learning itself.

6. Conclusion and implication

6.1. Conclusion

Based on the five-chain evidence of this three-group quasi-experiment, this study draws four conclusions.

First, a "non-outsourceable" agentic workflow can significantly change the surface form of outsourcing (direct dumping fell from 75.0% to 11.1%), but total volume remained unchanged (p = 0.864), persisting in more concealed, collaborative forms ("cosmetic modification", "division of labor"): "form substitution", not "behavioral correction".

Second, the strong-control workflow induced high-intensity "performative engagement": substantial time maneuvering around the system (25.4 minutes at Node 3) but markedly insufficient authenticity (comment matching 3.62; 27.8% of difficulty tags highly authentic). The time served "satisfying the system", not "making the brain understand".

Third, the workflow's detection logic constitutes a "hidden curriculum": over four months, students systematically learned to analyze detection rules, calibrate performance dosage, and organize collective evasion; every explicit "success" added bricks to tacit educational failure.

Fourth, only 1 of 120 students (0.8%) exhibited a self-regulated strategy, and metacognitive calibration was low across groups (high-calibration shares 9.1%–13.9%): as a habitus, "brain outsourcing" far exceeds the corrective capacity of a single-semester technical intervention.

6.2. Theoretical contribution

This study makes three theoretical contributions. First, the "hidden-curriculum paradox" extends Jackson's concept into agentic workflows, showing that technical governance turns "evading detection" into a course no one teaches yet everyone must take. Second, the "surface closed loop of governance" describes how a governance system and the governed form a self-sufficient exchange around behavioral indicators that hollows out the educational goal—a new empirical form of Foucauldian governmentality in educational-technology settings. Third, the applicable scope of desirable-difficulty theory is extended: imposed externally by algorithms without learners' metacognitive participation, difficulty becomes something to be evaded and performed, so the laboratory findings of cognitive psychology cannot be transplanted directly into technology-governed classrooms.

6.3. Practical implication

For instructional designers, the most direct implication is that "dialogue beats confrontation": Group B's negotiative workflow matched Group A on transfer (34.7 vs. 34.4, p = 0.989) at a much lower perceived-control cost (+0.96 vs. +2.17). Preserving the possibility of outsourcing while raising its metacognitive cost thus balances learning outcomes and student experience better than comprehensive blocking. Concretely, turning AI use from a "prohibited item" into an "auditable item"—requiring students to take a position on AI suggestions (agree/disagree/why)—is more likely than enforced "independent completion" to induce genuine cognitive processing.

For educational administrators, this study cautions against equating the intensity of technical surveillance with teaching quality. The nonsignificant between-group difference in the composite outsourcing index (p =

0.864) shows that behaviorally "good-looking data" can mask cognitive hollowness. Evaluations of technical interventions must include authenticity indicators and transfer indicators, rather than relying solely on completion rates and dwell time.

For policymakers, this study calls for attention to the educational-equity issue of "antecedent habitus". The "brain outsourcing" habitus of students at low-tier private undergraduate colleges is a sediment of prolonged educational-selection experience, not a simple moral defect or technology abuse. Any governance scheme that fails to address the structural causes behind this habitus can only work on the surface.

6.4. Limitations and future directions

This study has four limitations. First, internal validity: disciplinary background and treatment were fully confounded (Electronic Information → Group A; Artificial Intelligence → Group B); baseline checks (ability pre-test $F \approx 0.00$; SES $p = 0.203$) partially mitigate this, but unobserved variables such as disciplinary culture may still confound the results. Second, external validity: a single site, course, and semester limit transferability. Third, measurement: the authenticity of B4 difficulty tags and E3 calibration rests on coder judgment and retains some subjectivity despite double-coding checks ($\kappa = 0.72$). Fourth, ethics and data: no dedicated ethics approval was obtained because the institution does not mandate review for teaching research (protective measures were nonetheless implemented), and interview quotations are limited by the subsample size (10 per group).

Future research can proceed in three directions: cross-institutional and cross-course replication to test the generality of the "hidden-curriculum paradox"; longitudinal tracking of whether evasive habitus rebounds after the intervention ends; and combined designs of "metacognitive scaffolding" and negotiative workflows to test whether self-regulated strategies can be cultivated at scale without sacrificing transfer outcomes.

6.5. Concluding remark

When "brain outsourcing" becomes habitus, the dilemma of technical governance lies not in its being insufficiently refined but in its being too refined—refined enough to train every student into an expert at evading detection, yet never able to make them the master of their own learning. The governmentality paradox of agentic workflows reminds educators that the hardest problem in education has never been a technical problem but a human one; not a problem of detection but of trust; not how to let the system "see" learning, but how to let learners believe once again—that learning is worth doing in person.

Funding

The "Three Campuses" Teaching Reform Research Project, under the project "Research on Innovation of a Hybrid Online–Offline Teaching Model for Data Structures Based on Collaboration between AI Teaching Assistants and the Offline Classroom" (2025SGXYJG014)

References

[1] Bearman, M., Ryan, J., & Ajjawi, R. (2022). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education, 86*(2), 369-385. https://doi.org/10.1007/s10734-022-00937-2

[2] Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, 28. https://doi.org/10.1186/s40561-024-00316-7

[3] Prather, J., Reeves, B. N., Leinonen, J., MacNeil, S., Randrianasolo, A. S., Becker, B. A., Kimmel, B., Wright, J., & Briggs, B. (2024). The widening gap: The benefits and harms of generative AI for novice programmers. In *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1* (pp. 469–486). Association for Computing Machinery. https://doi.org/10.1145/3632620.3671116

[4] Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Article 680, pp. 1–24). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642902

[5] Zhai, X. (2024). *Do AI tutors empower or enslave learners? Toward a critical use of AI in education*. arXiv. https://doi.org/10.48550/arXiv.2507.06878

[6] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676-688. https://doi.org/10.1016/j.tics.2016.07.002

[7] Zhang, S., Zhao, X., Zhou, T., & Zhang, S. (2024). Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. *International Journal of Educational Technology in Higher Education, 21*, 34. https://doi.org/10.1186/s41239-024-00467-0

[8] Prather, J., Reeves, B. N., Denny, P., Becker, B. A., Leinonen, J., Luxton-Reilly, A., Powell, G., Finnie-Ansley, J., & Santos, E. A. (2023). "It's weird that it knows what I want": Usability and interactions with Copilot for novice programmers. *ACM Transactions on Computer-Human Interaction, 31*(1), Article 4. https://doi.org/10.1145/3617367

[9] Prather, J., Leinonen, J., Kiesler, N., Benario, J. G., Lau, S., MacNeil, S., Norouzi, N., Opel, S., Pettit, V., Porter, L., Reeves, B. N., Savelka, J., Smith, D. H., Strickroth, S., & Zingaro, D. (2025). Beyond the hype: A comprehensive review of current trends in generative AI research, teaching practices, and tools. In *2024 Working Group Reports on Innovation and Technology in Computer Science Education* (pp. 300–338). Association for Computing Machinery. https://doi.org/10.1145/3689187.3709614

[10] Reicherts, L., Zhang, Z. T., von Oswald, E., Liu, Y., Rogers, Y., & Hassib, M. (2025). AI, help me think—but for myself: Assisting people in complex decision-making by providing different kinds of cognitive support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Article 255, pp. 1–19). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713295

[11] Beattie, L. (2020). Educational leadership: Producing docile bodies? A Foucauldian perspective on higher education. *Higher Education Quarterly, 74*(1), 98–110. https://doi.org/10.1111/hequ.12218

[12] Rovea, F. (2020). Intercultural education and everyday life: Suggestions from Michel de Certeau. *Ethics and Education, 15*(1), 62–76. https://doi.org/10.1080/17449642.2019.1669878

[13] Rovea, F. (2021). Teaching as altered knowledge: Rethinking the teaching practice with Michel de Certeau. *Studies in Philosophy and Education, 41*(1), 1-15. https://doi.org/10.1007/s11217-021-09792-7

[14] Mu, M. G., & Dooley, K. (2023). *Recontextualising and recontesting Bourdieu in Chinese education: Habitus, mobility and language*. Routledge. https://doi.org/10.4324/9781003327301

[15] Espinoza, O., González, L., Sandoval, L., & McGinn, N. (2023). The relationship between class-based habitus and choice of university and field of study. *British Journal of Sociology of Education, 44*(4), 649-668. https://doi.org/10.1080/01425692.2023.2194563

[16] Heather, M. (2023). Reproduction to transformation: Disrupting teacher habitus through pedagogic work. *British Journal of Sociology of Education, 44*(2), 355-373. https://doi.org/10.1080/01425692.2022.2161473

[17] Meyer, R., Archer, E., & Smit, L. (2025). The positive influence of the hidden curriculum in medical education: A scoping review. *Medical Science Educator, 35*(3), 1-10. https://doi.org/10.1007/s40670-025-02380-1

[18] Nina, R., & Liezel, F. (2023). A conceptual framework for uncovering the hidden curriculum in private higher education. *Cogent Education, 10*(1). https://doi.org/10.1080/2331186X.2023.2191409

[19] Gallagher, M., Breines, M., & Blaney, M. (2020). Ontological transparency, (in)visibility, and hidden curricula: Critical pedagogy amidst contentious edtech. *Postdigital Science and Education, 3*(2), 1-19. https://doi.org/10.1007/s42438-020-00198-1

[20] Naya, J. (2023). Reimagining Freire: Beyond human relations. *Cultural Studies of Science Education, 18*(1), 205-216. https://doi.org/10.1007/s11422-023-10154-7

[21] Peuler, M., & Harlow, S. (2025). Critical pedagogies in online instructional design within academic libraries: A scoping review. *The Journal of Academic Librarianship, 51*(5), 103116. https://doi.org/10.1016/j.acalib.2025.103116

[22] Bjork, E. L., & Bjork, R. A. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

[23] Ab, H. B., B., F., H., L., & B., A. (2023). Worth the effort: The start and stick to desirable difficulties (S2D2) framework. *Educational Psychology Review, 35*(2). https://doi.org/10.1007/s10648-023-09766-w

[24] Rutherford, S. (2020). Using desirable difficulty concepts to enhance durable learning in design education. *Art, Design & Communication in Higher Education, 19*(1), 65-79. https://doi.org/10.1386/adch_00014_1

[25] National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. National Academies Press. https://doi.org/10.17226/9853

[26] Marko, T. (2022). Education and technology: Key issues and debates. *International Review of Education, 68*(4), 635-636. https://doi.org/10.1007/s11159-022-09971-9

[27] Segneda, A., Kone, P. E., & Hongde, L. (2025). Massification in higher education: A systematic review of its boundaries, drivers, and the role of critical pedagogy. *International Journal of Higher Education, 14*(6). https://doi.org/10.5430/ijhe.v14n6p108

[28] The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8. https://doi.org/10.3102/0013189X032001005

[29] Khizar, U. M. H., Iqbal, J. M., Khalid, J., & Iqbal, S. (2020). Student impression management and academic performance: A moderated mediation model. *Journal of Public Affairs, 21*(3). https://doi.org/10.1002/pa.2258