2025 Volume 3, Issue 7
Educational Innovation Research
ISSN: 3029-1844(Print)
ISSN: 3029-1852(Online)

Vibe Learning: Cultivating Mathematical Modelling Literacy in High School: A Quasi-Experimental Mixed-Methods Study

Zhihao Wang[1], Dechen Fan[2], Tingting Niu[1]
[1]Shanghai Jingye High School, Shanghai, 200000, China
[2]Zhejiang University, Zhejiang, 310027, China

Abstract

Vibe learning brings the idea-first ethos of vibe programming into classroom practice: learners set intentions and evaluation criteria while technology recedes. We report a quasi-experimental mixed-methods study in two parallel Grade 11 classes: an experimental class used an AI-augmented modelling sandbox with a Socratic micro-tutor across three scenarios, while a control class received traditional instruction. Data included baseline questionnaires, platform logs, post-task micro-surveys, and summative assessments. Vibe learning enabled full-cycle modelling within regular periods: completion time fell by 33%, first-attempt correctness rose, submission cycles dropped, and feedback latency shrank from minutes to seconds. Both classes reached similar concept mastery by end of the term, with larger self-efficacy gains in the experimental group. We outlined orchestration to balance efficiency with productive struggle and situate the design in pragmatism, social constructivism, phenomenology, hermeneutics, and virtue epistemology, showing a scalable, human-centered path to modelling literacy with transparent, computable evidence.

Keywords

AI tutor; Mathematical modelling literacy; Vibe learning

Online publication: August 26, 2025

1. Introduction

Mathematical modelling literacy, problem framing, model construction, analysis, validation, and contextual decision making is essential yet difficult to enact within ordinary lessons. Time pressure, delayed feedback, and uneven process visibility often compress modelling into procedural exercises [1]. We present vibe learning, a design where human ideas lead and technology recedes. The teacher and students co-formulate intentions and criteria; a lightweight sandbox and Socratic AI micro-tutor support exploration, verification, and traceable revision.

Building on a school-built atmospheric learning platform, the research questions are as follows:

(1) Does vibe learning improve process efficiency and accuracy in modelling tasks?

(2) How does it affect learning outcomes and self-efficacy relative to traditional instruction?

(3) What trade-offs arise for cognitive load, initiative, and sense of accomplishment?

The contributions of this study are as listed:

(1) A classroom-tested vibe learning work flow with a minimal evidence chain (original → AI feedback → resubmission → teacher review);

(2) Empirical results from a quasi-experiment with mixed methods in authentic lessons;

(3) A philosophical account clarifying why and how idea-first, human-centered orchestration can coexist with powerful AI scaffolds.

2. Vibe learning framework

Vibe learning comprises three continuous stages:

(1) Vibe ideation (ideas first): From an authentic situation, learners clarify goals, variables, constraints, boundary conditions, and acceptance criteria without fixing a method in advance [2];

(2) Vibe programming (rapid prototyping): Natural-language specifications are translated into classroom-ready sandboxes (web apps) with AI assistance. Technical details stay invisible; learning mechanics and evaluation remain explicit [3];

(3) Vibe learning (implementation): Students plan first, then interact with the sandbox for parameter probing, sensitivity checks, and validation. A Socratic micro-tutor gives just-in-time hints and epistemic prompts (no answers). All attempts and revisions are logged as a minimal evidence chain to support formative assessment and reflection [4].

3. Methodology

3.1. Participants and setting

Two intact Grade 11 classes in a public model high school (Shanghai) participated (n = 80; experimental class, n = 40; control class, n = 40). The same teacher taught both classes; devices were available in the experimental class. Permissions and ethics approvals were secured [5].

3.2. Instruments and data sources

Baseline questionnaire was used to investigate demographics, prior exposure to modelling and AI tools, and a 13-item modelling self-efficacy scale (problem analysis, model construction, tool use, validation, communication; 1–5 Likert) [6]. Platform logs (experimental class) were also employed for timestamped actions (parameter edits, submissions), AI interactions, automated correctness checks, and derived metrics such as time-on-task, submission cycles, first-attempt correctness, and feedback latency [7]. Post-task micro-surveys were utilized to assess cognitive load (7-point mental effort), engagement and perceived usefulness. Finally, summative assessments, which are a paper-based modelling test (novel scenario, no AI) and a short concept inventory were used.

3.3. Intervention tasks and control condition

The intervention tasks for experimental classes and conditions for control classes are as outlined below:

(1) Quadratic inequality:

(i) Experimental class: Express constraints from a textual context as a quadratic inequality; visualize solution sets; verify with representative points;

(ii) Control class: parallel content on paper with teacher checks [8,9];

(2) Walking in the rain:

(i) Experimental class: Model total rain exposure as a function of speed (with wind); explore for optimal speed; test extremes and limitations;

(ii) Control class: algebraic reasoning and hand-drawn graphs [10];

(3) Financial option valuation:

(i) Experimental class: Parameterize a European call stylized model; explore sensitivity (e.g., volatility); validate reasonableness via edge cases;

(ii) Control class: lecture plus worksheet practice [11].

In all experimental classes, students first plan without hints (*"human brain first"*), then use the sandbox while control classes used conventional explanation, board work, and paper tasks [12].

3.4. Design and analysis

A quasi-experimental design compared classes on process and outcome measures. Descriptive statistics summarized effects; inferential tests guided interpretation where distributional assumptions held. Interviews (sampled students) and observations provided qualitative context via thematic analysis [13].

4. Results

4.1. Baseline equivalence

Both classes were comparable at baseline (gender balance, recent math performance, prior exposure to modelling and AI tools) (**Table 1**).

Table 1. Baseline background (experimental vs. control, n = 80)

Characteristic	Experimental	Control
Avg. math percentile (%)	50	52
Modelling exposure: none/ occasional/systematic (%)	48/45/7	50/40/10
AI tools: never/occasional/regular (%)	60/35/5	55/40/5

4.2. Process efficiency and accuracy

Across tasks, the experimental class was faster and more accurate, with fewer cycles and markedly shorter feedback loops [14] (**Table 2**).

Table 2. Process and performance metrics (task averages)

Metric	Experimental	Control
Completion time (relative)	≈ -30%	Baseline
Advanced task time (T3)	≈ -50%	Baseline
First-attempt correctness	≈ 65%	≈ 40%
Final success rate	≈ 95%	≈ 70%
Submission cycles (median)	3	≈ 5

4.3. Learning outcomes and self-efficacy

Both classes achieved strong end-of-term modelling performance on a novel paper task (no AI). The experimental class showed a higher mean (trend) and larger self-efficacy gains [15].

4.4. Cognitive load and perceptions

Control students reported higher mental effort (~ +1.0 to +1.2 on a 7-point scale) and more time awaiting teacher feedback (**Table 3**). Students of experimental classes praised immediacy and reduced frustration; a minority expressed concern about over-relying on hints and welcomed structured “think-first” windows [16].

Table 3. Selected pre-post measures

Measure	Experimental		Control	
	Pre	Post	Pre	Post
Modelling self-efficacy (1–5)	2.6	3.8	2.5	3.5
Modelling test (0–100)	55	85	56	79
Concept inventory (0–20)	-	16.5	-	16.2

5. Discussion

5.1. Advantages of vibe learning

Vibe learning offers several advantages as follows:

(1) Feasible full-cycle modelling: Near-real-time guidance makes open-ended tasks executable within a single period, including advanced content typically deferred [17];

(2) Computable formative evidence: The minimal evidence chain supports granular feedback, targeted re-teaching, and reflective debriefs; teachers gain visibility into process, not just product;

(3) Comparability and reuse: Uniform sandboxes standardize affordances across classes, enabling cross-cohort comparison and iterative improvement [18].

5.2. Tensions and orchestration

Faster convergence and higher first-try accuracy can reduce trial diversity and perseverance opportunities. Hence, we recommend a phased choreography:

(1) Sense-making first (no hints, timed): Representations, assumptions, plan [19];

(2) Targeted prompts to nudge strategy (not supply steps);

(3) Verification-without-answers: Tools confirm/falsify; students explain discrepancies [20];

(4) Meta-reflection: A short "what changed my mind" note. Lower extraneous load increases throughput but may blunt the "earned" feeling [21];

(5) Preserve productive difficulty with delayed hints and epistemic questions (e.g., "What would count as evidence against your model?").

5.3. Philosophical grounding

Pragmatism frames modelling as disciplined inquiry: hypothesize, intervene (simulate), observe consequences, revise. The sandbox is a laboratory for warranted assertions rather than answer-getting [22].

Social constructivism situates knowledge in mediated dialogue among teacher, peers, and AI; the micro-tutor functions as a mediator, not an oracle, to maintain learner agency [23]. Phenomenology opens problem-worlds (rain, traffic, markets) so structures-to-be-mathematized become present-to-experience [24]. Hermeneutics underscores iterative interpretation across symbol and context; students loop between formalism and narrative, refining meaning [25]. Virtue epistemology and phronesis (practical wisdom) guide when not to help: cultivate curiosity, perseverance, and intellectual humility through designed, bounded struggle [26].

6. Limitations and ethics

This single-site quasi-experiment used intact classes and a single instructor; broader generalization requires multi-site replication. Some measures (e.g., self-efficacy) rely on self-report. Data were minimized, anonymized, and locally archived; the classrooms used emphasized transparent disclosure of AI roles and assistance [27].

7. Conclusion

Vibe learning operationalizes idea-first, evidence-rich pedagogy for modelling. In authentic lessons, it delivered faster, more accurate task completion with lower cognitive burden, while maintaining end-of-term concept mastery and boosting self-efficacy. To avoid eroding independence, orchestration should protect human sense-making before assistance and require reflective justification after it. As schools seek scalable paths to modelling literacy, vibe learning offers a human-centered design where technology supports, but does not eclipse judgment.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Koedinger K, Corbett A, Perfetti C, 2012, The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. Cognitive Science, 36(5): 757–798.

[2] VanLehn K, 2011, The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46(4): 197–221.

[3] Shute V, 2008, Focus on Formative Feedback. Review of Educational Research, 78(1), 153–189.

[4] Kirschner P, Sweller J, Clark R, 2006, Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist, 41(2): 75–86.

[5] Sweller J, 2011, Cognitive Load Theory. Psychology of Learning and Motivation, 55.

[6] Kasneci E, Sessler K, Küchemann S, et al., 2023, ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. Learning and Individual Differences, 2023(103): 102274.

[7] Zawacki-Richter O, Marín V, Bond M, et al., 2019, Systematic Review of Research on Artificial Intelligence Applications in Higher Education: Where Are The Educators? International Journal of Educational Technology in Higher Education, 16(1): 39.

[8] Graesser A, Person N, Magliano J, 2004, AutoTutor: A Tutor with Dialogue in Natural Language. Behavior Research Methods, Instruments, & Computers, 36(2): 180–192.

[9] D'Mello S, Graesser A, 2012, AutoTutor and Affective AutoTutor: Learning by Talking with Cognitively and Emotionally Intelligent Computers that Talk Back. ACM Transactions on Interactive Intelligent Systems, 2(4): 1–39.

[10] Mislevy R, Almond R, Lukas J, 2004, A Brief Introduction to Evidence-Centered Design. Tech Rep, 2004: 632.

[11] Mislevy R, 2005, Evidence-Centered Assessment Design: Layers, Structures, and Terminology. PADI Tech Rep, 2005: 9.

[12] Luckin R, Holmes W, Griffiths M, et al., 2016, Intelligence Unleashed: An Argument for AI in Education, Pearson, London.

[13] Blum W, Galbraith P, Henn H, et al., 2007, Modelling and Applications in Mathematics Education: The 14th ICMI Study, Springer, New York.

[14] Blum W, Ferri R, 2009, Mathematical Modelling: Can It be Taught and Learnt? Journal of Mathematical Modelling and Application, 1(1): 45–58.

[15] Kaiser G, Blum W, Ferri R, et al., 2011, Trends in Teaching and Learning of Mathematical Modelling, Springer, The Netherlands.

[16] Webster F, 1958, Traffic Signal Settings, Road Research Tech, Her Majesty's Stationery Office, London.

[17] Lighthill M, Whitham G, 1955, On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads. Proceedings of the Royal Society A, 229(1178): 317–345.

[18] Richards P, 1956, Shock Waves on the Highway. Operations Research, 4(1): 42–51.

[19] Bocci F, 2012, Whether or Not to Run in the Rain. European Journal of Physics, 33(5): 1321–1332.

[20] Usher E, Pajares F, 2009, Sources of Self-Efficacy in Mathematics: A Validation Study. Contemporary Educational Psychology, 34(1): 89–101.

[21] Betz N, Hackett G, 1983, The Relationship of Mathematics Self-Efficacy Expectations to the Selection of Science-Based College Majors. Journal of Vocational Behavior, 23(3): 329–345.

[22] Black F, Scholes M, 1973, The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3): 637–654.

[23] Merton R, 1973, Theory of Rational Option Pricing. The Bell Journal of Economics and Management Science, 4(1): 141–183.

[24] Dewey J, 1938, Logic: The Theory of Inquiry, Henry Holt, New York.

[25] Vygotsky L, 1978, Mind in Society: The Development of Higher Psychological Processes, Harvard Univ Press, Cambridge.

[26] Gadamer H, 2004, Truth and Method, 2nd rev ed, Bloomsbury, London.

[27] Aristotle, 1999, Nicomachean Ethics, 2nd ed, Hackett, Indianapolis.

Publisher's note

Whioce Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.