Teaching Gradient Descent

An Exploratory Study on Classic Textbook vs. Multiple Representations Approaches

Fabiana-Maria Severin

Delft University of Technology, EEMCS

Supervisors: Gosia Migut and Ilinca Rențea · Examiner: Jorge Martinez Castaneda

CSE3000 Research Project – June 2026

Learning materials

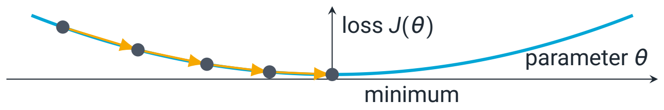
Analysis code

1. Why this matters

Gradient descent is compact as a formula, but difficult as a mental model. Beginners must coordinate mathematical notation, a repeated procedure, and an intuition for movement on a loss surface.

formula · graph · iteration · intuition

$$\theta_{t+1} = \theta_t - \alpha \nabla J(\theta_t)$$



A learner may reproduce the update rule while still struggling to explain the loss, gradient, learning rate, negative-gradient direction, or repeated updates.

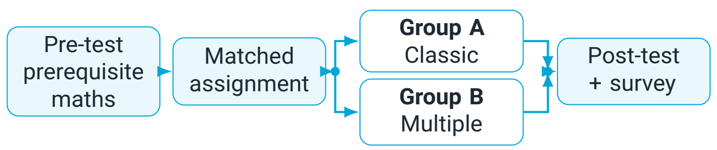
2. Research questions

RQ1 How does learning gradient descent with multiple representations affect students' conceptual understanding and problem solving performance compared with a classic textbook-style method?

RQ2 How does learning gradient descent with multiple representations affect students' perceived clarity, confidence, cognitive load, usefulness, and engagement compared with a classic textbook-style method?

3. Methodology

Design. Controlled exploratory study with **15 first-year CS students**; matched assignment to two learning conditions.



Group A: Classic textbook

- Text and definitions
- Mathematical notation
- Static visual support
- Short examples

Group B: Multiple representations

- Geometric visualizations
- Worked example
- Downhill analogy
- Interactive exploration

Same learning goals, core content, study time, and assessment

Expert evaluation before data collection. A machine learning professor reviewed the materials for accuracy, comparability, alignment, and cognitive-load risks. Five revisions followed:

concepts first · worked-example steps · analogy mapping · interactive alignment · clearer wording

4. Mixed-methods analysis

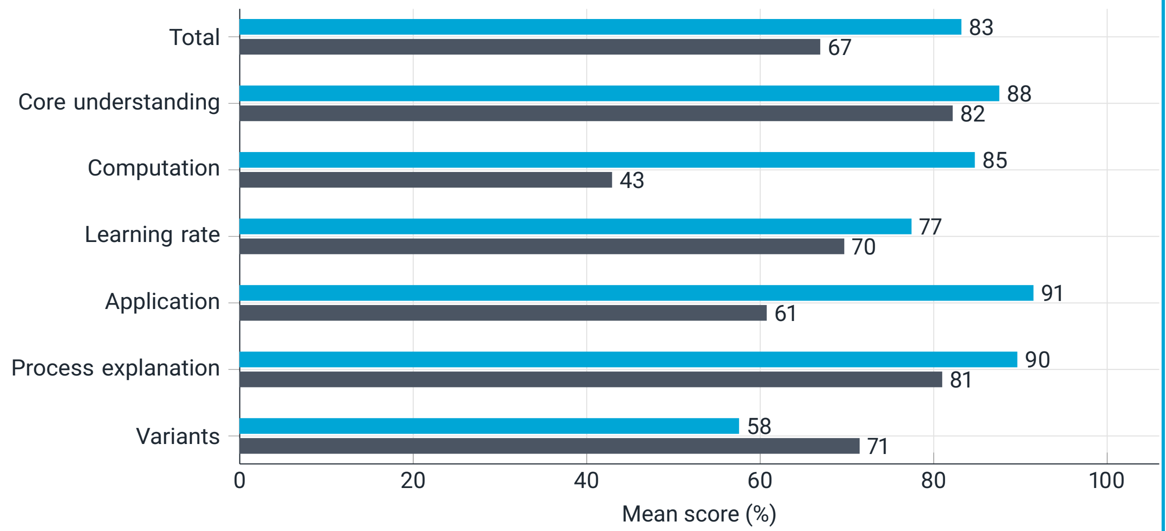
Quantitative analysis

- Post-test total and six sub-scores
- Six learner-experience constructs
- Mann–Whitney U, Cliff's δ (exploratory; $N = 15$)

Qualitative analysis

- Think-aloud notes + open comments
- Hybrid thematic coding; participant code counts
- Reliability: 85% agreement, $\kappa = .70$

5. Learning outcomes: strongest gains in computation and application



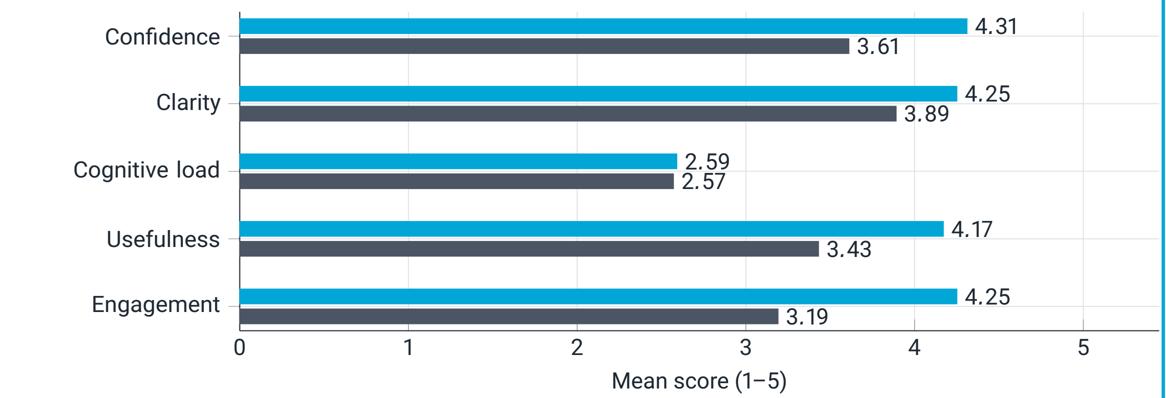
+16.3 total points $p = .006$, $\delta = .86$

9.4 → 18.6 computation /22 +9.20 points

9.7 → 14.6 application /16 +4.92 points

7.1 → 5.8 variants /10 −1.39 points

6. Learner experience: more engaging, not more demanding



Engagement +1.06 $p = .023$, $\delta = .71$

Usefulness +0.74 $\delta = .59$

Load +0.02 essentially unchanged

7. Qualitative findings: explaining the quantitative pattern through code frequency

Computation +9.20 points
- Worked example used: 4/8 multiple
- Worked example requested: 3/7 classic

Application +4.92 points
- Visual support used: 5/8 multiple
- Interactive support used: 4/8 multiple
- Visual support requested: 4/7 classic

Variants -1.39 points
- Variant confusion: 5/8 multiple vs. 2/7 classic

Learner experience
- Extraneous load: 2/7 classic vs. 0/8 multiple

8. Implications for teaching gradient descent

- Worked examples are especially useful for beginners.
- Visual and interactive supports help with transfer.
- Match each representation to the reasoning required by the task.
- Intuitive supports should be linked to formal terminology.
- More representations are not always better; alignment matters.

9. Limitations & future work

- **Small exploratory sample:** $N = 15$ limits generalizability.
- **Immediate measurement:** long-term retention was not assessed.
- **Combined intervention:** individual representation effects cannot be isolated.
- **Next steps:** replicate with a larger sample, delayed post-tests, and separate representation combinations.

10. Conclusion

In this exploratory sample, aligned multiple representations supported both **learning performance** and **learner experience**. The clearest benefits appeared in **computation**, **application**, **usefulness**, and **engagement**, while perceived cognitive load remained almost unchanged.

Selected references: Rențea, Migut, and Krijthe (2025), *Are Interactive Visualizations in Machine Learning Education Helping Students?*; Atkinson et al. (2000), *Learning from Examples: Instructional Principles from the Worked Examples Research*; Mayer (2005), *Cognitive Theory of Multimedia Learning*; Sweller (1988), *Cognitive Load During Problem Solving: Effects on Learning*.

TU Delft Delft University of Technology