When the Pattern Comes Too Easily: What an AI Simulation Takes Away in the Optics Lab

Xingyu Shu*[1]

[1]The High School Attached to Zhejiang University, Hangzhou, China

Abstract

Generative AI now allows teachers to build clean, parameter-driven optics simulations in minutes, making them tempting substitutes for time-consuming laboratory setups. In this classroom comparison, 96 Grade 11 students studied Young's double-slit interference through two routes: one group worked with physical apparatus, while the other used an AI-generated web simulation and spent the freed-up time on derivation and practice. The AI-simulated group performed significantly better on the same-day written assessment (86.6 vs. 69.3 overall; $p < 0.001$), especially on calculation items (90.6 vs. 62.1; Cohen's $d = 2.58$). Yet when both groups later faced an unscripted single-slit diffraction task, the pattern reversed: more than half of the AI-simulated pairs (54.2%) fell into repetitive refinement or scale-misreading behaviors, while most physical-apparatus pairs (87.5%) showed exploratory iteration or extended investigation. The case suggests that AI can improve test readiness, but if simulations remove too much experimental friction, they may also remove the empirical judgment students need for real inquiry.

Study Design

Participants

Ninety-six Grade 11 students from regular classes were divided into two groups of 48 by class. The school forms its regular classes to be balanced in mean achievement, score variance, gender ratio, and other characteristics, and all of these students were taught physics by the author. Across the most recent unit tests the two groups' average scores differed by under 2%, so we treat them as equivalent at the start.

Conditions

The two groups studied the same idea—Young's double-slit interference—but encountered it through different classroom settings.

As shown in Fig. 1, the physical-apparatus group used an optical bench to align the beam, observe fringes, and calculate fringe spacing. The AI-simulated group used an AI-generated web simulation of the same experiment, adjusting wavelength λ, slit separation d, and screen distance L. The contrast was not between "old" and "new" instruction: Young's double-slit experiment can itself support student-centered inquiry,[1] and interactive simulations have a strong record in physics teaching when used thoughtfully.[2] Table 1 summarizes the resulting time trade-off.

*Correspondence: sh.u@my.cityu.edu.hk

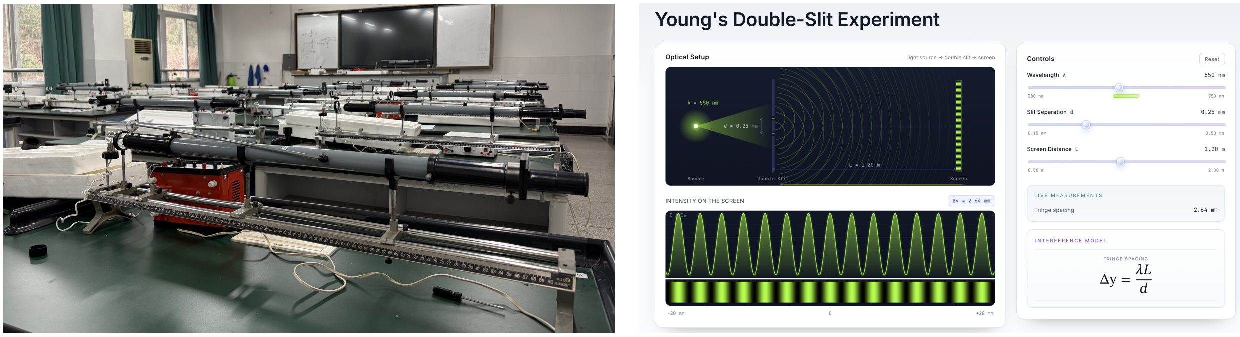


Figure 1: Two routes into Young's double-slit interference: the physical apparatus (left) and an AI-generated simulation (right).

Table 1: Time cost of the two routes.

Time category	Physical-apparatus group	AI-simulated group
Instructor / lab technician preparation	About 4 hours of setup and 2 hours of teardown by a lab technician	About 10 minutes of AI authoring; the web page is reusable across sections and years
Student class time	40 minutes at the bench: alignment, screen-distance adjustment, observation, recording, and calculation	10 minutes of simulation work; the remaining 30 minutes are used for derivation and problem practice

Much of the physical-apparatus group's class time went into alignment and imperfect observation. Figure 2 shows two typical results: a faint filtered-light pattern that counted as a success, and a white-light pattern produced when students forgot the color filter. Such mishaps are common in the lab, but the phenomena they reveal are not visible in the AI simulation.

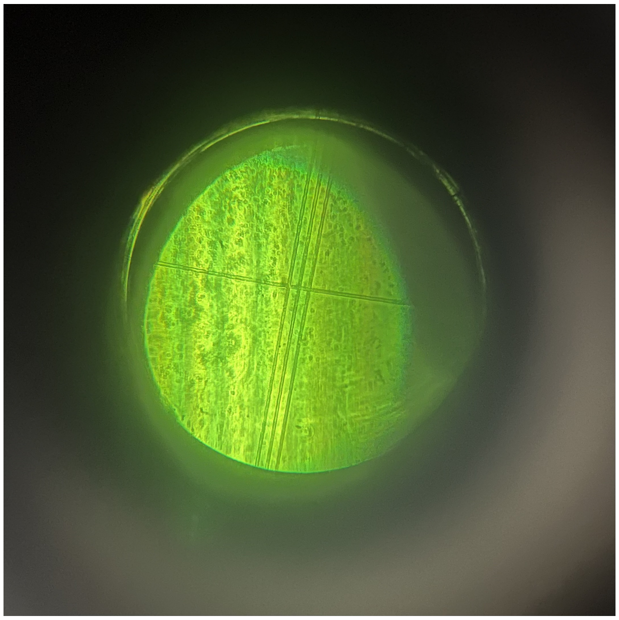
Written Assessment: Double-Slit Interference

On the same day as the lesson, every student took the same supervised written assessment. It combined multiple-choice items on interference concepts with calculation items requiring students to solve for λ from fringe spacing, slit separation, and screen distance.

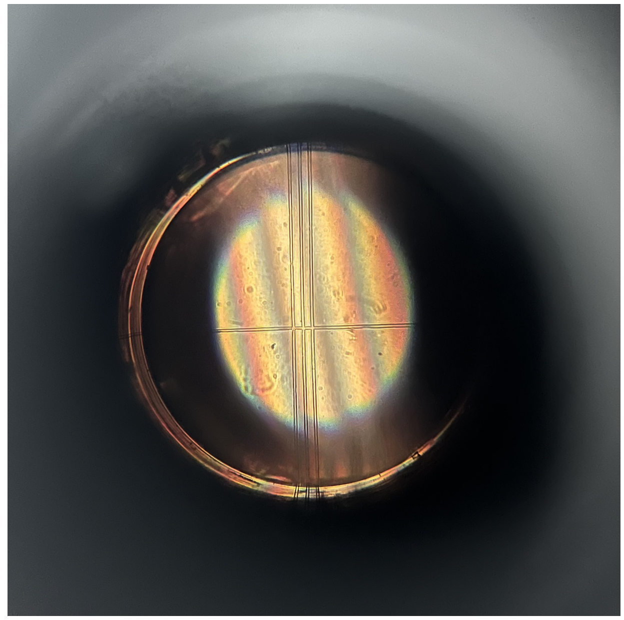
Table 2: Assessment scores (out of 100) by group and question type, mean ± SD ($n = 48$ per group).

Group ($n = 48$ students)	Overall	Multiple-choice	Calculation
Physical-apparatus	69.3 ± 14.5	74.6 ± 15.2	62.1 ± 14.4
AI-simulated	86.6 ± 5.7	84.5 ± 6.4	90.6 ± 6.1
p	< 0.001	< 0.001	< 0.001
Cohen's d	1.57	0.85	2.58

The AI-simulated group's higher scores (Table 2) likely reflect a time-on-task effect. This is



(a)



(b)

Figure 2: Student-produced interference from the physical apparatus: (a) a typical (successful) filtered-light result; (b) a white-light pattern produced when the color filter was omitted.

consistent with reviews showing that virtual and remote laboratories can produce strong outcomes on conventional assessments.[3] Here the simulation rendered fringe spacing as a direct, parameter-driven outcome, so students reached $\Delta y \approx \lambda L/d$ quickly and reinvested the saved time in worked examples, practice, and consolidation. The physical-apparatus group spent the full period at the bench and reached the formula with almost no time left for practice. The larger SD in that group may also reflect that some students had not yet produced an observable pattern by the end of class.

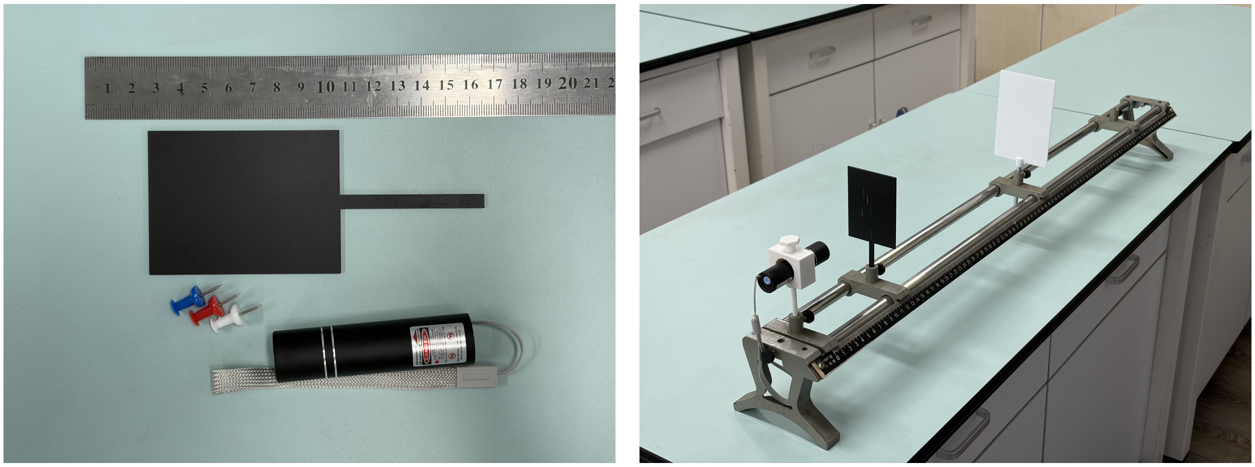
Open-Ended Task: Single-Slit Diffraction

Working in pairs, students were asked to design an experiment to observe diffraction. Each pair received only the items in Fig. 3(a): a steel ruler, a matte-black spray-painted acrylic plate with a handle, pushpins, and a low-power Class 2 visible laser pointer ($\leq$1 mW). The laser holder and screen were already mounted on the optical rail, but students had to build and align the optical path in Fig. 3(b). No step-by-step instructions were given; students decided how to scrape the slit, set its width, place it relative to the laser, choose the screen distance, and judge whether the result counted as diffraction. The same task was given to both groups.

Table 3 gives a concise behavior code for each category, and reports the percentage of pairs assigned to it. The visual panels of Fig. 4 show representative slit-making outcomes; the table describes the behavior behind them. The point of these categories is not only whether students produced a clean pattern, but how they compared uncertain results, revised procedures, and made experimental judgments—habits closely tied to critical thinking in laboratory instruction.[4]

As shown in Fig. 4 and Table 3, the outcomes fall into four categories.

Category (a), exploratory iteration, was the most common. After an unsatisfactory trial, these students cut a new slit and realigned the optical path. Their eventual success came through repeated attempts.



(a) **(b)**

Figure 3: The open-ended single-slit diffraction task: (a) the provided materials; (b) the optical path students needed to assemble.

Table 3: Categories of student behavior. Each pair was assigned to the category its final result most closely matched.

Category	Behavior	Physical	AI-sim.	Total
(a)	Exploratory iteration: restarted with new slits and realigned the optical path	27.1%	12.5%	39.6%
(b)	Extended investigation: went beyond the assigned task to test new aperture or wavelength conditions	16.7%	10.4%	27.1%
(c)	Repetitive refinement: kept the setup fixed and repeated the same operation	4.2%	18.8%	22.9%
(d)	Scale misreading: mistook enlarged schematic diagrams for the real scale	2.1%	8.3%	10.4%
Total	—	50.0%	50.0%	100.0%

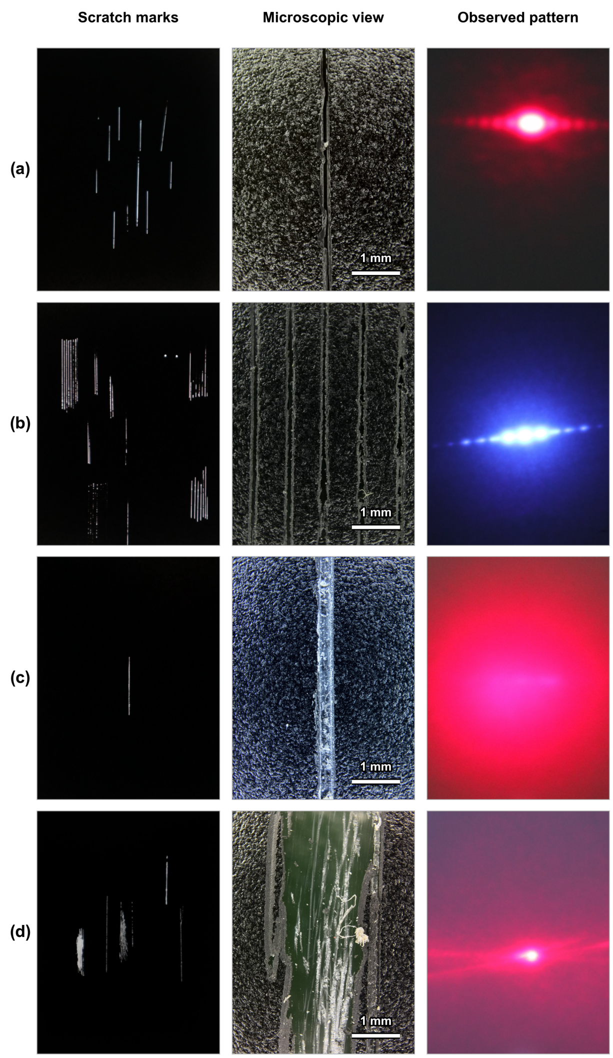


Figure 4: Student-made slit plates and their optical consequences. Rows (a)–(d) show representative outcomes arranged by scratch marks, microscopic view, and observed pattern. Scale bars in the microscopic views indicate 1 mm.

Category (b), extended investigation, covers pairs who continued after completing the assigned task. They used the apparatus to test double slits, gratings, circular apertures, hair, and pen tips. Some even held the laser pointer by hand to project the pattern onto a distant wall and asked for the lights to be turned off. They also requested laser pointers of different colors to check how wavelength affects the pattern. These students treated the setup as a tool for asking new questions.

Category (c), repetitive refinement, describes pairs who fixed the setup and tried to perfect one slit. Some observed an imperfect pattern but kept scratching the same position until the acrylic plate was damaged. Their problem was not a lack of effort, but an overly idealized expectation of how the experiment should look.

Category (d), scale misreading, involved deliberately widening the slit, apparently prompted by enlarged schematic diagrams. These students concluded that a single scratch was too narrow for light to pass through, and their report diagrams showed correspondingly wide slits. Pairs who later abandoned this approach and succeeded were counted in category (a).

Read against the written-assessment results, these behaviors invert the verdict of the written test. Once a pattern had to be produced rather than read off a slider, 54.2% of the AI-simulated group's pairs fell into categories (c) and (d), while the physical-apparatus group showed nearly the mirror-image distribution, with 87.5% of its pairs in categories (a) and (b).

Conclusion

This study presents a pedagogical paradox: the AI-simulated group outperformed on the written test but struggled when the same physics had to be transferred to an unstructured bench task. That contrast fits work showing that physics labs can build experimentation practices not captured by traditional content assessments.[5] In this case, the AI simulation compressed a noisy empirical process into a high-success, parameter-driven result. That design bought higher scores through this time-on-task advantage, but it also removed chances to correct, iterate, and tolerate failure when working with an uncooperative apparatus. The physical-apparatus group spent time on beam alignment that looked "wasted" but functioned as a cognitive investment in empirical judgment.

This result does not mean AI tools should be kept out of the physics laboratory. Physical and virtual manipulation can both support learning, but they do not necessarily provide the same experience.[6] The ethical design question is whether the technology supplements the empirical encounter or replaces it. In our lesson on reading a vernier caliper, students held a real caliper while the teacher used an AI tool to operate and enlarge a virtual caliper of the same model on screen. There the AI removed a visual obstacle that carried little conceptual value; it deepened the learning rather than standing between the student and the evidence.

Generative AI will inevitably become woven into physics teaching, but laboratory time has goals beyond content coverage, including experimental design, modeling, technical skill, data analysis, and communication.[7] Recent transformations of introductory labs make the same point: authentic practice matters.[8] When an AI simulation strips the anomalies out of an experiment and trades empirical ability for test readiness, efficiency can be misread as instructional progress. Responsible AI use in education requires transparency about what the tool makes visible and what it hides.[9] The real value of AI in the physics laboratory lies not in how much tedium it removes, but in whether it helps students connect more honestly with physical reality.

Ethics and AI Use Statement

The activities described here were part of the regular optics curriculum, so the study introduced no additional instructional burden or risk. Equity was preserved through a crossover design: after the assessment and open-ended task, the two groups exchanged conditions and ultimately received the same learning experiences. All scores, reports, and photographs were anonymized before analysis and are reported only in aggregate. An AI tool was used to generate the classroom simulation, and Claude Opus 4.8 was used to polish the English of this manuscript; the content was written entirely by the author, who has reviewed and verified all of it. The classroom design, data analysis, and final claims remain the author's responsibility.

References

[1] D. Z. Meyer, "A Student-Centered, Inquiry-Based Approach to Young's Double-Slit Experiment (and Other Investigations of Light's Wave Character)," Phys. Teach. **55**(3), 159–163 (2017).

[2] C. Wieman, W. K. Adams, P. Loeblein, and K. K. Perkins, "Teaching Physics Using PhET Simulations," Phys. Teach. **48**(4), 225–227 (2010).

[3] J. R. Brinson, "Learning outcome achievement in non-traditional (virtual and remote) versus traditional (hands-on) laboratories: A review of the empirical research," Comput. Educ. **87**, 218–237 (2015).

[4] N. G. Holmes, C. E. Wieman, and D. A. Bonn, "Teaching critical thinking," Proc. Natl. Acad. Sci. U.S.A. **112**(36), 11199–11204 (2015).

[5] E. M. Smith, M. M. Stein, C. Walsh, and N. G. Holmes, "Direct Measurement of the Impact of Teaching Experimentation in Physics Labs," Phys. Rev. X **10**, 011029 (2020).

[6] Z. C. Zacharia and G. Olympiou, "Physical versus virtual manipulative experimentation in physics learning," Learn. Instr. **21**(3), 317–331 (2011).

[7] J. Kozminski *et al.*, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (American Association of Physics Teachers, 2014), `https://www.aapt.org/resources/upload/LabGuidlinesDocument_EBendorsed_nov10.pdf`.

[8] S. F. Wolf and M. W. Sprague, "Introductory Physics Labs: A Tale of Two Transformations," Phys. Teach. **60**(5), 372–375 (2022).

[9] F. Miao and W. Holmes, *Guidance for Generative AI in Education and Research* (UNESCO, Paris, 2023), `https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research`.