

Trust but Verify

AI Verification Across the STEM Pipeline

Mike Borowczak¹ · Andrea C. Borowczak²

¹ Electrical & Computer Engineering, DRACO Lab · ² School of Teacher Education — University of Central Florida

mike.borowczak@ucf.edu · andrea.borowczak@ucf.edu · sciforum.net/event/IOCES2026

START HERE

follow the gold route 1 → 9

2.45 ≈ 2.48

KEY FINDING. Students rating AI-policy clarity (1-5) where AI is allowed without guidance, and educators rating their confidence verifying AI (1-5) before training, both sit near 2.5 / 5 — the same gap on both sides of the pipeline.

1

INTRODUCTION & AIM

The Question

Can students and teachers verify what AI produces?

RQ1: which course AI-policy types give students the least clarity?

RQ2: can a verification-focused workshop raise educators' confidence in verifying AI?

How we measured it ▶

2

METHOD

Two streams, one pipeline

Students: 128 course experiences (71 courses, 9 semesters) from 30 ECE students, plus a 147-student survey; clarity 1-5.

Educators: 33 K-12 teachers, matched pre/post (n=24), six 1-5 skills.

Tests: ANOVA, Welch, paired t.

Into the clarity gap ▼

4

THE SHADOW CURRICULUM

Used anyway

⚠ 62.5%

of students used AI in courses that banned it (15/24) — unguided.

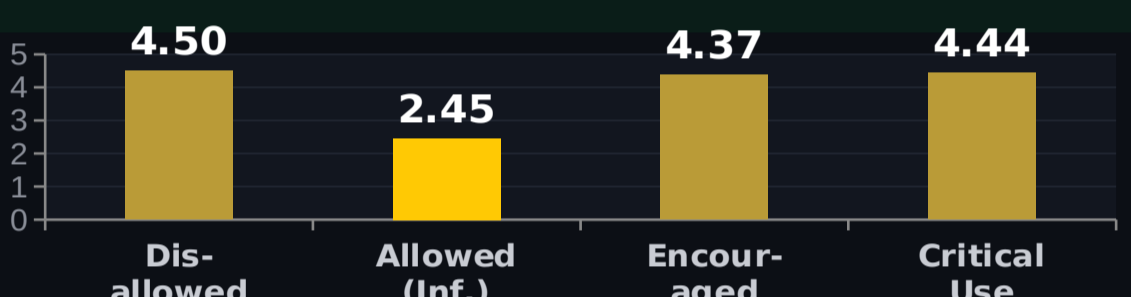
Top tasks: explanation 45.8%, ideation 33.3%, debugging 29.2%.

Find the road out ▼

3

STUDENTS

The Clarity Gap



◀ They used it anyway

5

THE ROADMAP

Depth of Integration

Figure 1

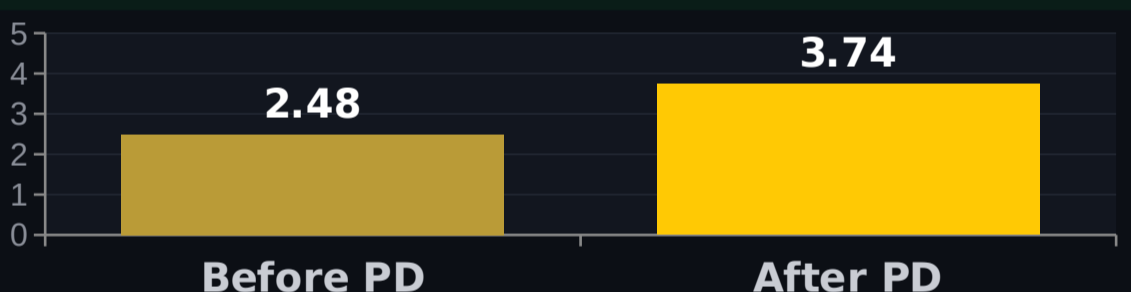


Now, the educators ▶

6

K-12 EDUCATORS

Verification Self-Efficacy



See the biggest lift ▼

8

CONCLUSION

What closes the gap

Policy silence yields the lowest clarity — below a ban. Verification is teachable: one workshop produced large, uniform gains.

Build verification-first structure — Critical Use for students, verification PD for educators.

Explore related work → DRACO Lab · CRAFT studies

Read every number ▼

7

CHECK THE MACHINE

The biggest lift



1.65 → 3.35

Check the Machine confidence (1-5): before → after, +1.70 — the largest gain.

◀ What closes the gap

9

MEASURES & FORMULAS

How to read every number

Scale: all ratings 1-5 Likert (1 low, 5 high).

M / SD: mean rating and its spread.

Composite: mean of the six skill items.

%: share of respondents (62.5% = 15/24).

Cohen's d: $(M_1 - M_2) / \text{pooled SD}$ — group effect.

d_z: mean change ÷ SD of change — paired effect.

F / p: ANOVA variance test; chance probability.

Table 1. Educator self-efficacy, matched pre/post (1-5, n=24)

Verification skill	Pre	Post	Δ
Composite	2.48	3.74	+1.26
Identify AI errors	2.70	3.70	+1.00
Verify AI code	2.21	3.21	+1.00
Check the Machine ★	1.65	3.35	+1.70
Teach evaluation	2.33	3.83	+1.50
Cross-check sources	2.50	3.79	+1.29

REFERENCES & ACKNOWLEDGMENTS

Borowczak & Borowczak ('26). From Allowed to Embedded; Trust but Verify. ASEE/ASEE-SE. GLSVLSI '25 · Educ. Sci. 9(1).

Educator PD supported in part by NSF #2230997 (WRNN); views are the authors' own.

More from the authors:

DRACO Lab · M. Borowczak (Scholar) · A. Borowczak (Scholar)

AUTHORS

Mike Borowczak, PhD

Assoc. Professor, ECE · Director, DRACO Lab · UCF

Andrea C. Borowczak, PhD

Professor & Director, School of Teacher Education

· UCF