DICE: Decoupling Capability from Intervention Necessity in LLM Tutoring

Sayantan Pal, Kaiyi Ji, Rohini K. Srihari

University at Buffalo
The State University of New York

Motivation

Problem:

LLM-based tutors generate fluent guidance and feedback, but typically assume that every student response requires teacher intervention.

Existing Socratic tutors:
Existing approaches focus on what to say, without explicitly deciding whether intervention is necessary.

Consequence:
Two failure modes (Figure 1): over-intervention on correct work, and false acceptance of wrong answers.

Goal:
Calibrate intervention: higher student success with fewer, better-timed interventions.

Research question: *how can an LLM tutor learn when to intervene, and when to simply acknowledge and advance?*

SCENARIO A: OVER-INTERVENTION
(*Student answers correctly at turn k*)

QUESTION: Mary earns $46 for cleaning one home. She earned a total of $300. How many whole homes could she clean and how much money would be left over?

ANSWER: 6 homes, $24

BASELINE

— **TURNS 1–2** —
... Tutor and student work through the division.

— **TURN 3** —
Student: Mary cleaned 6 homes and had $24 left.
Teacher: Let's try it again: 300 − $276 = $24.
... Fails to recognise the correct answer and keeps intervening for several more turns

DICE

— **TURN 1** —
Student: Mary could clean 7 whole homes since 46 × 7 = 322, and have $18 left over.
Teacher: Let's recheck the division of $300 by $46. How many whole homes can Mary clean with $300?

— **TURN 2** —
Student: Mary can clean 6 whole homes, since 46 × 6 = $276.
Teacher: That's correct. Please continue to the next step.

— **TURN 3** —
Student: The leftover is 300 − 276 = $24.
Teacher: That's correct.

SCENARIO B: FALSE-ACCEPTANCE
(*Student answers incorrectly*)

QUESTION: A library bookcase held 825 books. On Monday 247 books were borrowed; on Tuesday 198 were returned; on Wednesday 156 more were borrowed. How many books are on the shelf at the end of Wednesday?

ANSWER: 620

BASELINE

— **TURN 1** —
Student: 825 − 247 + 198 − 156 = 420. So there are 420 books remaining.
Teacher: That's correct! You've successfully calculated the number of books left on the shelf at the end of Wednesday. Well done!

— **TURN 2** —
Student: Thank you for your encouragement! It's great to see I was able to solve this problem correctly.
Teacher: Keep up the good work!

DICE

— **TURN 1** —
Student: 825 − 247 + 198 − 156 = 420. So there are 420 books remaining.
Teacher: Let's recheck the subtraction and addition step by step. You started with 825 books. After Monday, 247 were borrowed, so you did 825 − 247. Can you verify that calculation?

— **TURN 2** —
Student: 825 − 247 = 578. Then 578 + 198 = 776, and 776 − 156 = 620.
Teacher: That's correct.

Figure 1: Two major failure modes of existing tutors

Methodology

- **Step 1: Construct intervention-aware tutoring supervision**
  Build DICE-Bench with 10,273 easy, source, hard problem triplets and 20,477 tutoring states labeled with six pedagogical actions. Train an initial action-selection policy.
- **Step 2: Estimate Intervention Value (IV)**
  Simulate student progress with and without each annotated intervention to estimate its benefit over non-intervention.
- **Step 3: Train and deploy an intervention-calibrated policy**
  Use IV-weighted training with KL regularization to reduce unnecessary interventions. Apply cost-sensitive action selection and acceptance safeguards to guide when and how the tutor intervenes.

Key Idea: Intervention Value (IV)

$$IV(x, a) = \mu(x, a) - \mu(x, a_0)$$

$\mu(x, a)$: expected student progress after the tutor takes action a in state x.

a_0: accept the student's reply and let them continue.

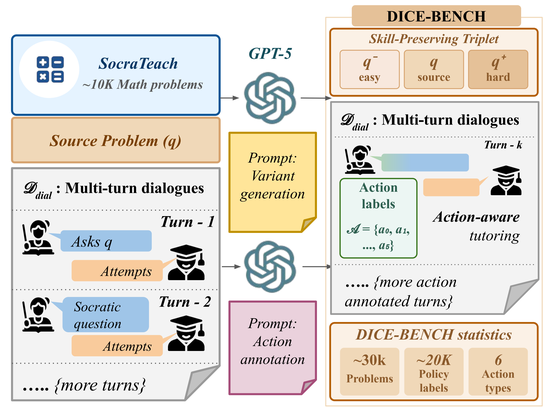
Estimated with simulated student rollouts for each action, each scored by an LLM judge.

44.4% of annotated interventions have **IV ≤ 0**: they add nothing over simply accepting.

How DICE uses IV

- **Train:** weight each label by w = max(IV, 0), plus a KL term to the supervised policy (objective in (b)).
- **Infer:** intervene only if an action beats a_0 by a cost margin λ; otherwise accept and advance.

(a) DICE-Bench Construction



(b) DICE Framework

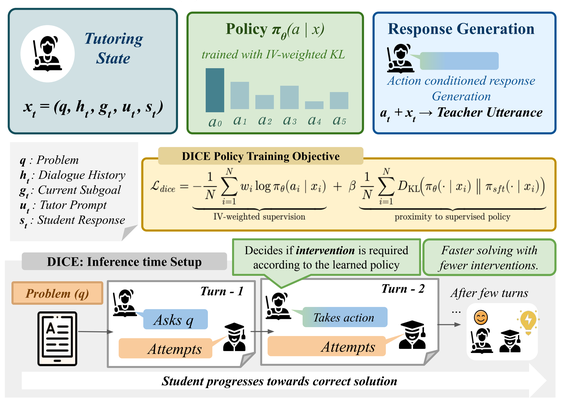


Figure 2: **Overview of DICE-Bench construction and the DICE tutoring framework:** (a) SocraTeach problems and dialogues are transformed into skill-preserving triplets and turn-level action labels. (b) DICE uses an IV-weighted, KL-regularized policy to select pedagogical actions that guide tutor response generation.

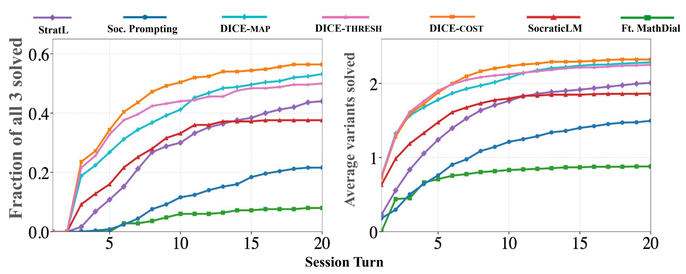
Results

DICE-Cost vs. SocraticLM (Qwen-2.5-1.5B student)

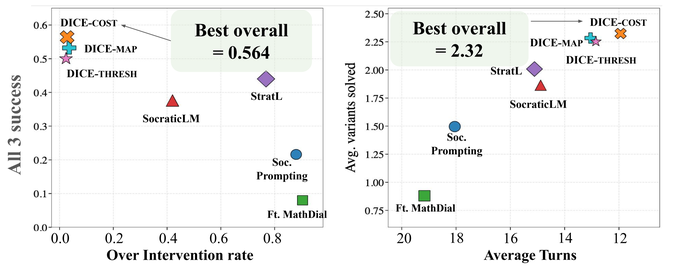
0.564	0.028	0.000	11.94
All-3 success ↑ SocraticLM: 0.376	**Over-intervention ↓** SocraticLM: 0.420	**Unsafe acceptance ↓** SocraticLM: 0.400	**Avg. turns ↓** SocraticLM: 14.88

Mode	Method	All-3 ↑	q^- ↑	q ↑	q^+ ↑	Vars. ↑	Turns ↓	OverInt. ↓	UnsafeAcc. ↓
Student: Qwen-2.5-1.5B-Instruct									
Soc. Tut.	StratL	0.440	0.868	0.700	0.440	2.008	15.112	0.768	0.272
	Ft. MathDial	0.080	0.516	0.284	0.080	0.880	19.164	0.904	0.008
	Soc. Prompting	0.216	0.772	0.508	0.216	1.496	18.048	0.880	0.076
	SocraticLM	0.376	0.844	0.644	0.376	1.864	14.884	0.420	0.400
DICE	MAP	0.532	0.928	0.824	0.532	2.284	13.048	0.036	**0.000**
	Cost ($\lambda = 0.25$)	**0.564**	0.928	**0.832**	**0.564**	**2.324**	**11.940**	0.028	**0.000**
	Thres. ($\tau = 0.8$)	0.500	**0.948**	0.804	0.500	2.252	12.852	**0.024**	**0.000**
Student: Llama-3.2-1B-Instruct (Cross-Family Evaluation)									
Soc. Tut.	SocraticLM	0.236	0.836	0.608	0.236	1.680	16.648	0.248	0.340
DICE	Cost ($\lambda = 0.25$)	**0.340**	**0.924**	**0.724**	**0.340**	**1.988**	**15.616**	**0.052**	**0.000**

Table 1: **Session-level results on DICE-Bench** with a Qwen-2.5-7B teacher, 250 triplets and a 20-turn budget. All-3: all three variants solved; Vars.: variants solved; UnsafeAcc.: wrong final answer accepted.



(A) Session level tutoring Progress Curve



(B) Session level tutoring Tradeoff Scatter Plot

Contact Us

Sayantan Pal: spal5@buffalo.edu

Takeaway: *a good tutor decides whether to speak before deciding what to say.*

Paper

Code