How to build tutoring systems that are almost as effective as human tutors?

Kurt VanLehn

School of Computing, Informatics and Decision Systems Engineering
Arizona State University

Outline

Next

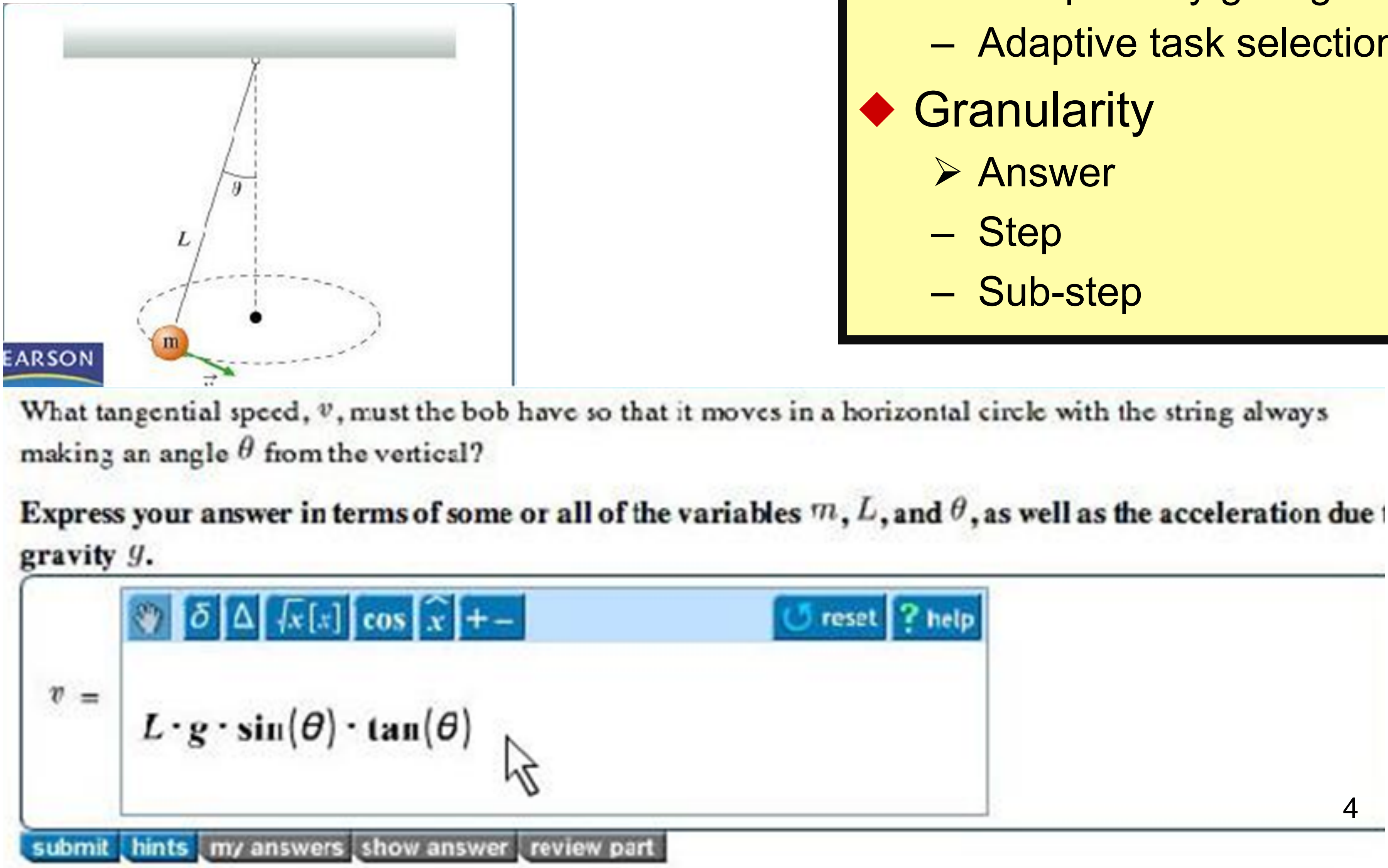
- Types of tutoring systems
- Step-based tutoring ≈ human tutoring
- How to build a step-based tutor
- Increasing their effectiveness
- Flame

Two major design dimensions

- Personalization of assignments
  - Non-adaptive
  - Competency gating
    - using sequestered assessments
    - one factor per module
  - Adaptive task selection
    - using embedded assessments
    - one factor per knowledge component
- Granularity of feedback, hints & other interaction
  - o Assignment (e.g., conventional homework)
  - Answer (e.g., most regular tutoring systems)
  - Step (e.g., most Intelligent Tutoring Systems)
  - Sub-step (e.g., human tutors & some ITS)

Example: Pearson's Mastering Physics

- ◆ Personalization
  - – Non-adaptive
  - ➢ Competency gating
  - – Adaptive task selection
- ◆ Granularity
  - ➢ Answer
  - – Step
  - – Sub-step



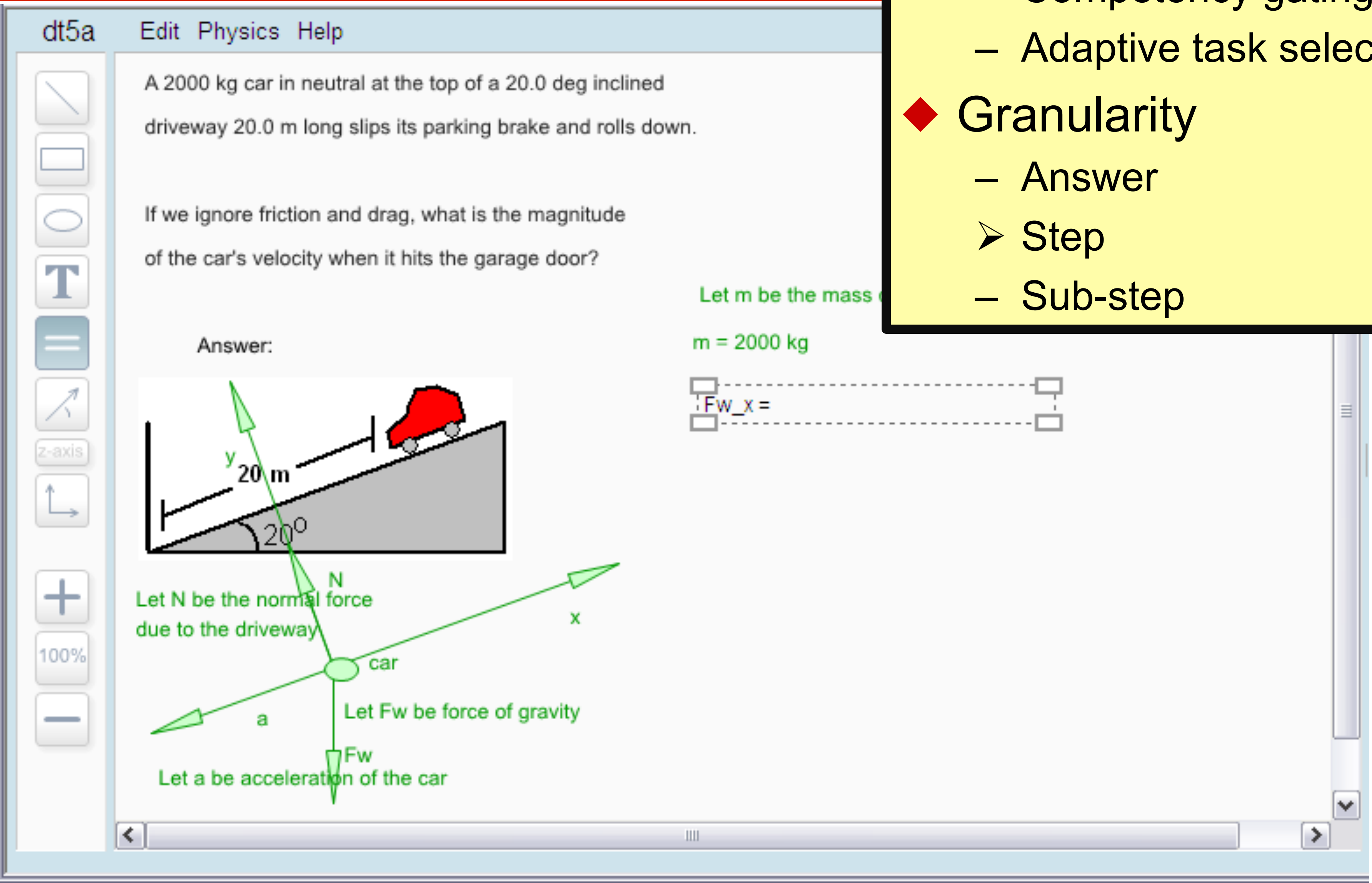
What tangential speed, v, must the bob have so that it moves in a horizontal circle with the string always making an angle θ from the vertical?

Express your answer in terms of some or all of the variables m, L, and θ, as well as the acceleration due to gravity g.

$v = L \cdot g \cdot \sin(\theta) \cdot \tan(\theta)$

Example: Andes Physics Tutor

- Personalization
  - ➢ Non-adaptive
  - – Competency gating
  - – Adaptive task selection
- Granularity
  - – Answer
  - ➢ Step
  - – Sub-step



A 2000 kg car in neutral at the top of a 20.0 deg inclined driveway 20.0 m long slips its parking brake and rolls down.

If we ignore friction and drag, what is the magnitude of the car's velocity when it hits the garage door?

Answer:

Let m be the mass

m = 2000 kg

Fw_x =

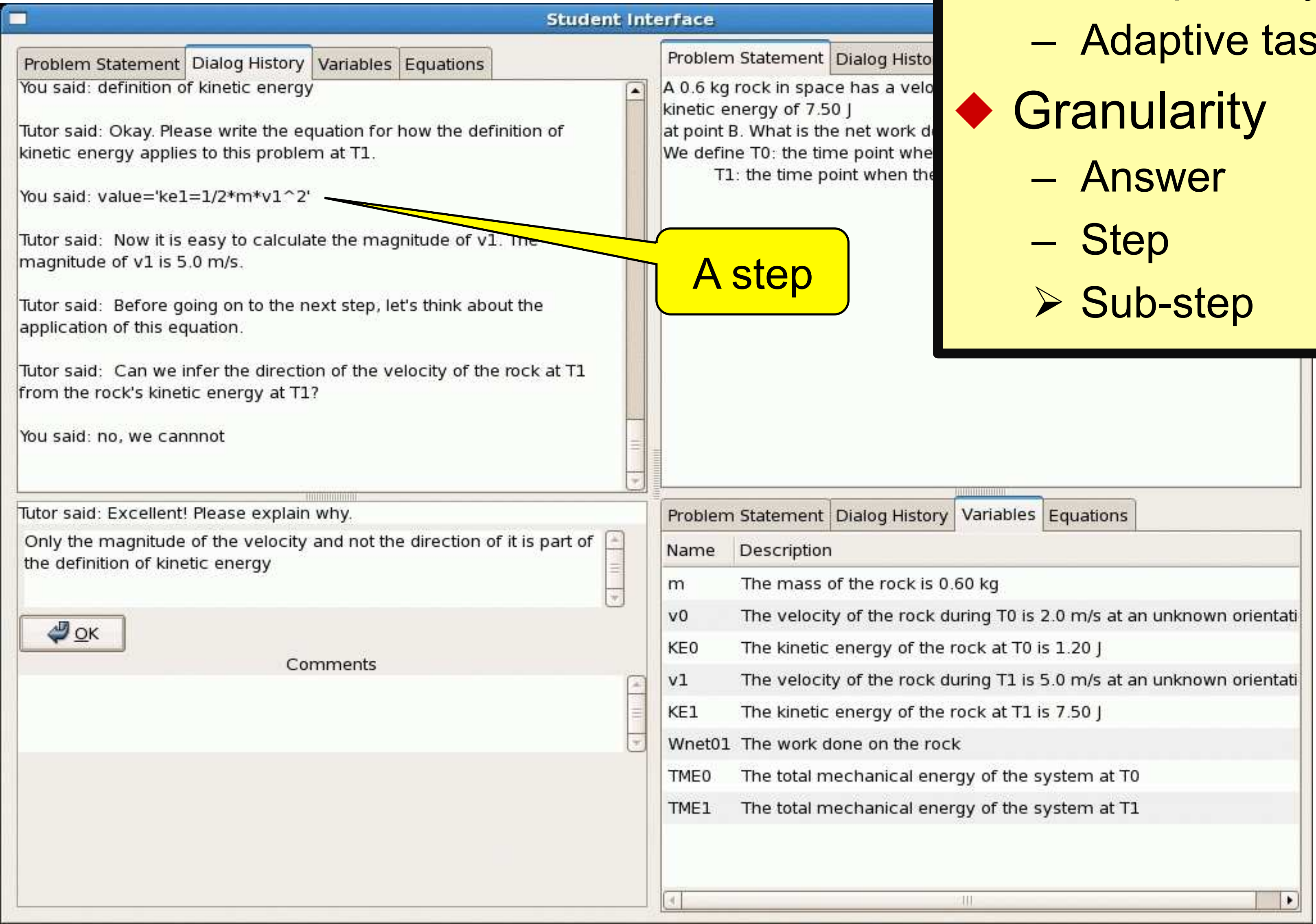
Let N be the normal force due to the driveway

Let Fw be force of gravity

Let a be acceleration of the car

Example: Cordillera Physics Tutor

- Personalization
  - Non-adaptive
  - – Competency gating
  - – Adaptive task selection
- Granularity
  - – Answer
  - – Step
  - Sub-step



You said: definition of kinetic energy

Tutor said: Okay. Please write the equation for how the definition of kinetic energy applies to this problem at T1.

You said: value='ke1=1/2*m*v1^2'

Tutor said: Now it is easy to calculate the magnitude of v1. The magnitude of v1 is 5.0 m/s.

Tutor said: Before going on to the next step, let's think about the application of this equation.

Tutor said: Can we infer the direction of the velocity of the rock at T1 from the rock's kinetic energy at T1?

You said: no, we cannnot

Tutor said: Excellent! Please explain why.

Only the magnitude of the velocity and not the direction of it is part of the definition of kinetic energy

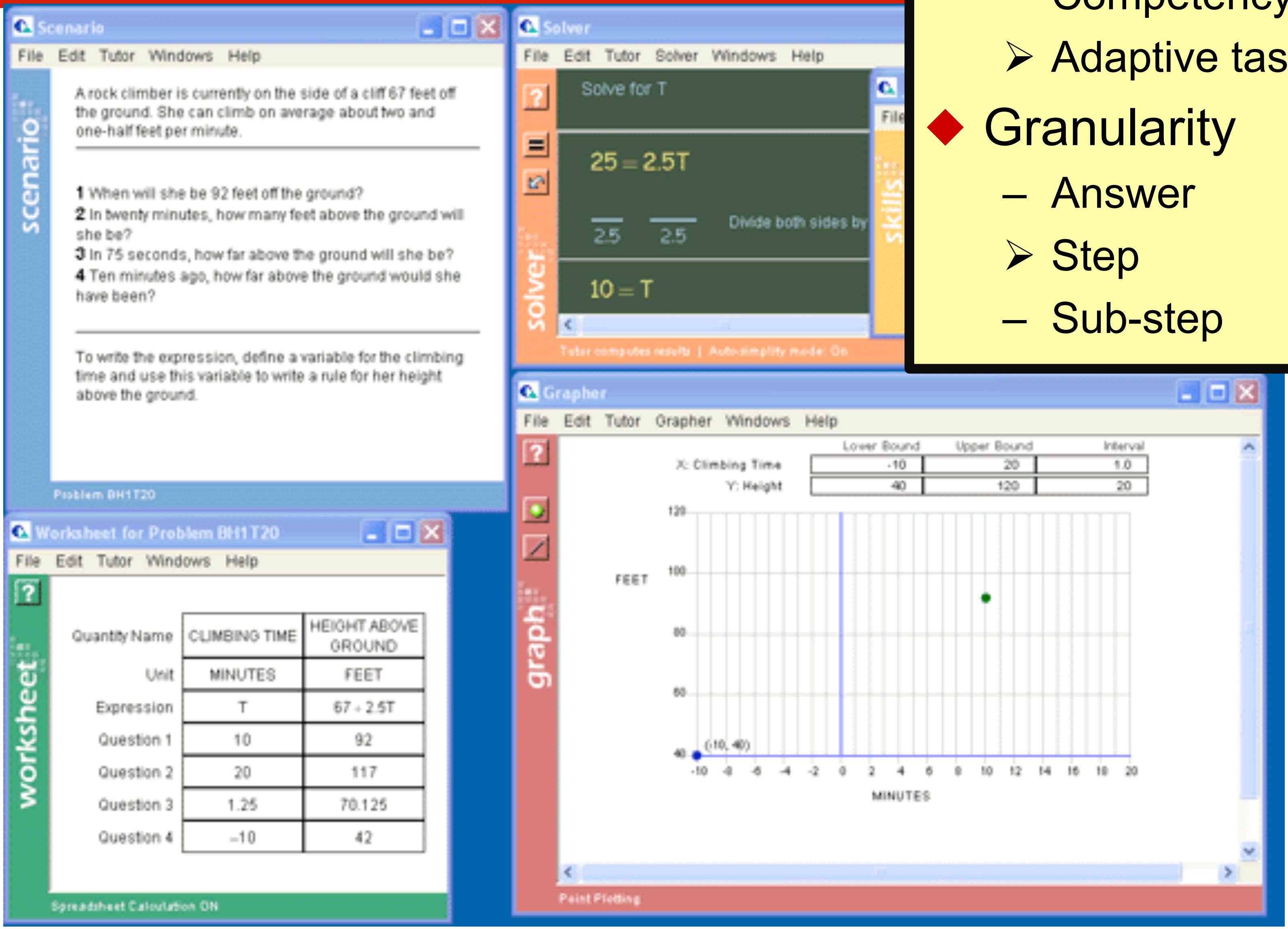
A 0.6 kg rock in space has a velo
kinetic energy of 7.50 J
at point B. What is the net work d
We define T0: the time point whe
T1: the time point when the

Name	Description
m	The mass of the rock is 0.60 kg
v0	The velocity of the rock during T0 is 2.0 m/s at an unknown orientati
KE0	The kinetic energy of the rock at T0 is 1.20 J
v1	The velocity of the rock during T1 is 5.0 m/s at an unknown orientati
KE1	The kinetic energy of the rock at T1 is 7.50 J
Wnet01	The work done on the rock
TME0	The total mechanical energy of the system at T0
TME1	The total mechanical energy of the system at T1

A step

Example: Carnegie Learning's Tutors

- Personalization
  - Non-adaptive
  - Competency gating
  - ➢ Adaptive task selection
- Granularity
  - Answer
  - ➢ Step
  - Sub-step


Scenario
File Edit Tutor Windows Help
scenario
Problem BH1T20
Solver
File Edit Tutor Solver Windows Help
solver
skills
Tutor computes results | Auto-simplify mode: On
Grapher
File Edit Tutor Grapher Windows Help
graph
Point Plotting
Worksheet for Problem BH1T20
File Edit Tutor Windows Help
worksheet
Spreadsheet Calculation ON
FEET
MINUTES
120
100
80
60
40
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
14
16
18
20
(-10, 40)

A rock climber is currently on the side of a cliff 67 feet off the ground. She can climb on average about two and one-half feet per minute.

1 When will she be 92 feet off the ground?
2 In twenty minutes, how many feet above the ground will she be?
3 In 75 seconds, how far above the ground will she be?
4 Ten minutes ago, how far above the ground would she have been?

To write the expression, define a variable for the climbing time and use this variable to write a rule for her height above the ground.

Solve for T

$25 = 2.5T$

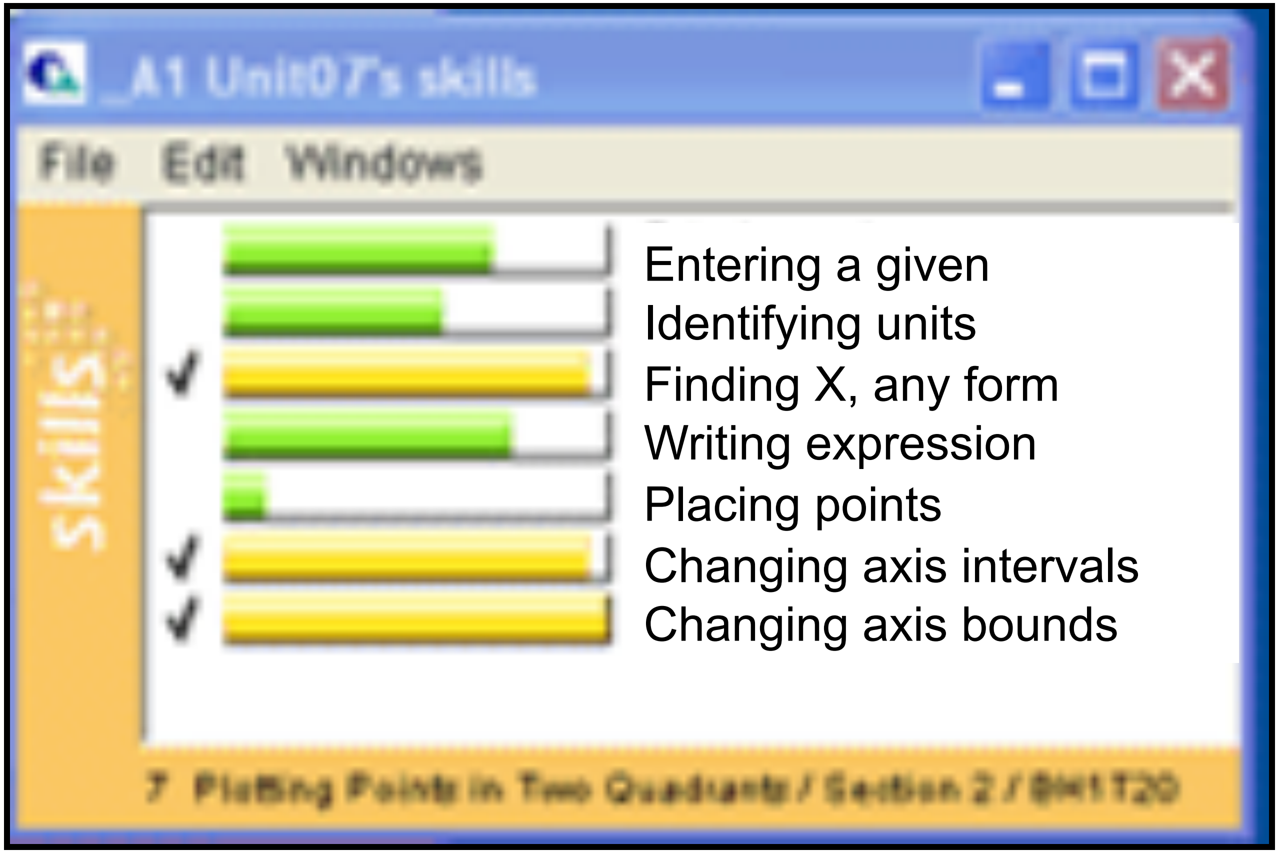
$\frac{}{2.5} \quad \frac{}{2.5}$ Divide both sides by

$10 = T$

Quantity Name	CLIMBING TIME	HEIGHT ABOVE GROUND
Unit	MINUTES	FEET
Expression	T	67 + 2.5T
Question 1	10	92
Question 2	20	117
Question 3	1.25	70.125
Question 4	-10	42

	Lower Bound	Upper Bound	Interval
X: Climbing Time	-10	20	1.0
Y: Height	40	120	20

Carnegie Learning's skillometer shows knowledge components & current competence



Entering a given

Identifying units

✓ Finding X, any form

Writing expression

Placing points

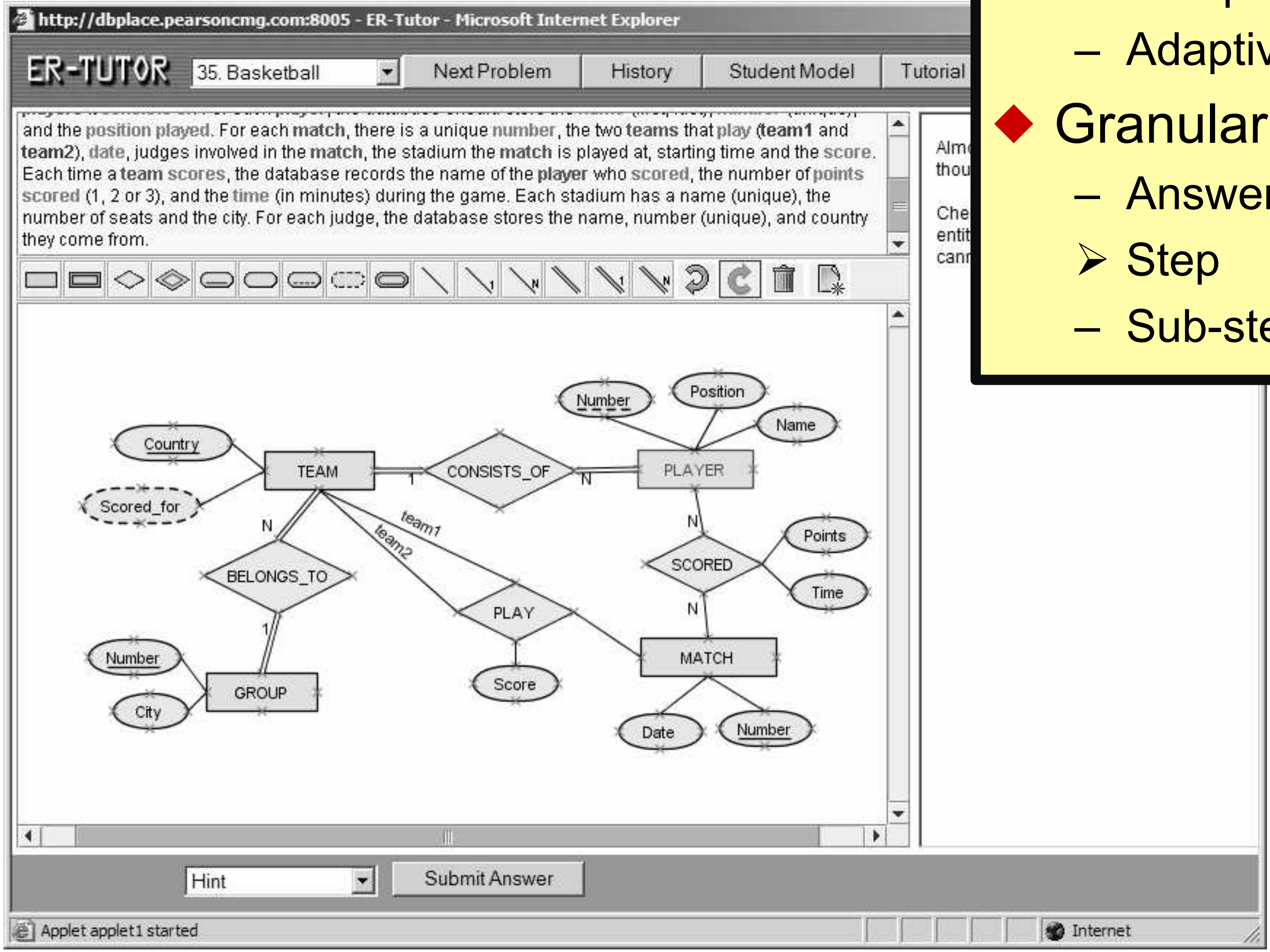
✓ Changing axis intervals

✓ Changing axis bounds

7 Plotting Points in Two Quadrants / Section 2 / BH1T20

Example: Entity-relation Tutor

- Personalization
  - ➢ Non-adaptive
  - – Competency gating
  - – Adaptive task selection
- Granularity
  - – Answer
  - ➢ Step
  - – Sub-step



http://dbplace.pearsoncmg.com:8005 - ER-Tutor - Microsoft Internet Explorer
ER-TUTOR
35. Basketball
Next Problem
History
Student Model
Tutorial

and the position played. For each match, there is a unique number, the two teams that play (team1 and team2), date, judges involved in the match, the stadium the match is played at, starting time and the score. Each time a team scores, the database records the name of the player who scored, the number of points scored (1, 2 or 3), and the time (in minutes) during the game. Each stadium has a name (unique), the number of seats and the city. For each judge, the database stores the name, number (unique), and country they come from.

Alm
thou

Che
entit
canr

Availability

	Non-adaptive	**Competency gating**	**Adaptive task selection**
Answer-based feedback/hints	Thousands	Hundreds	Few
Step-based feedback/hints	Hundreds (few on market)	Tens	Few
Sub-step based feedback/hints	Tens	None	None

Called CAI, CBT, CAL…

	Non-adaptive	**Competency gating**	**Adaptive task selection**
Answer-based feedback/hints	Thousands	Hundreds	Few
Step-based feedback/hints	Hundreds (few on market)	Tens	Few
Sub-step based feedback/hints	Tens	None	None

Called Intelligent Tutoring Systems (ITS)

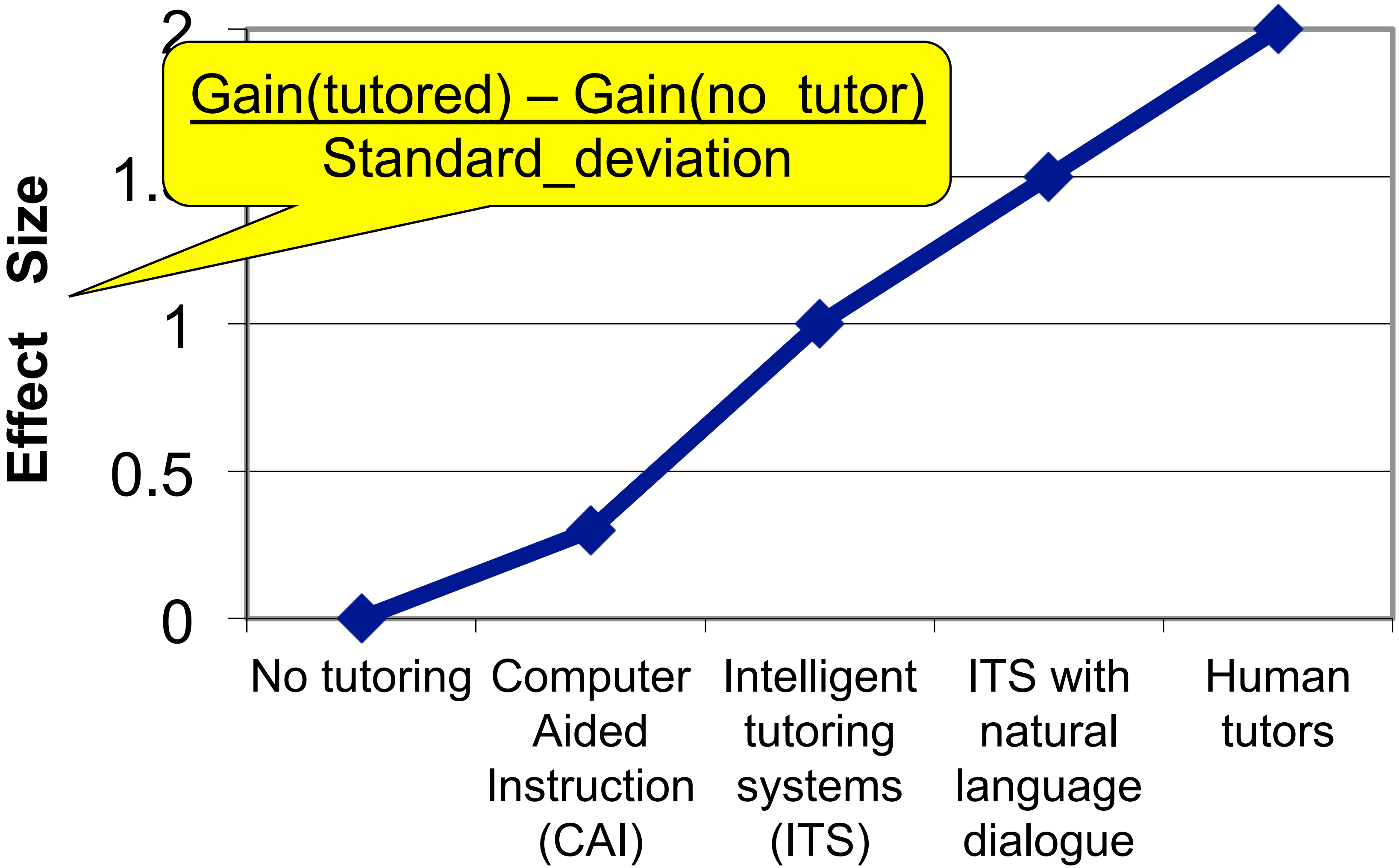
	Non-adaptive	Competency gating	Adaptive task selection
Answer-based feedback/hints	Thousands	Hundreds	Few
Step-based feedback/hints	Hundreds (few on market)	Tens	Few
Sub-step based feedback/hints	Tens	None	None

Outline

Next

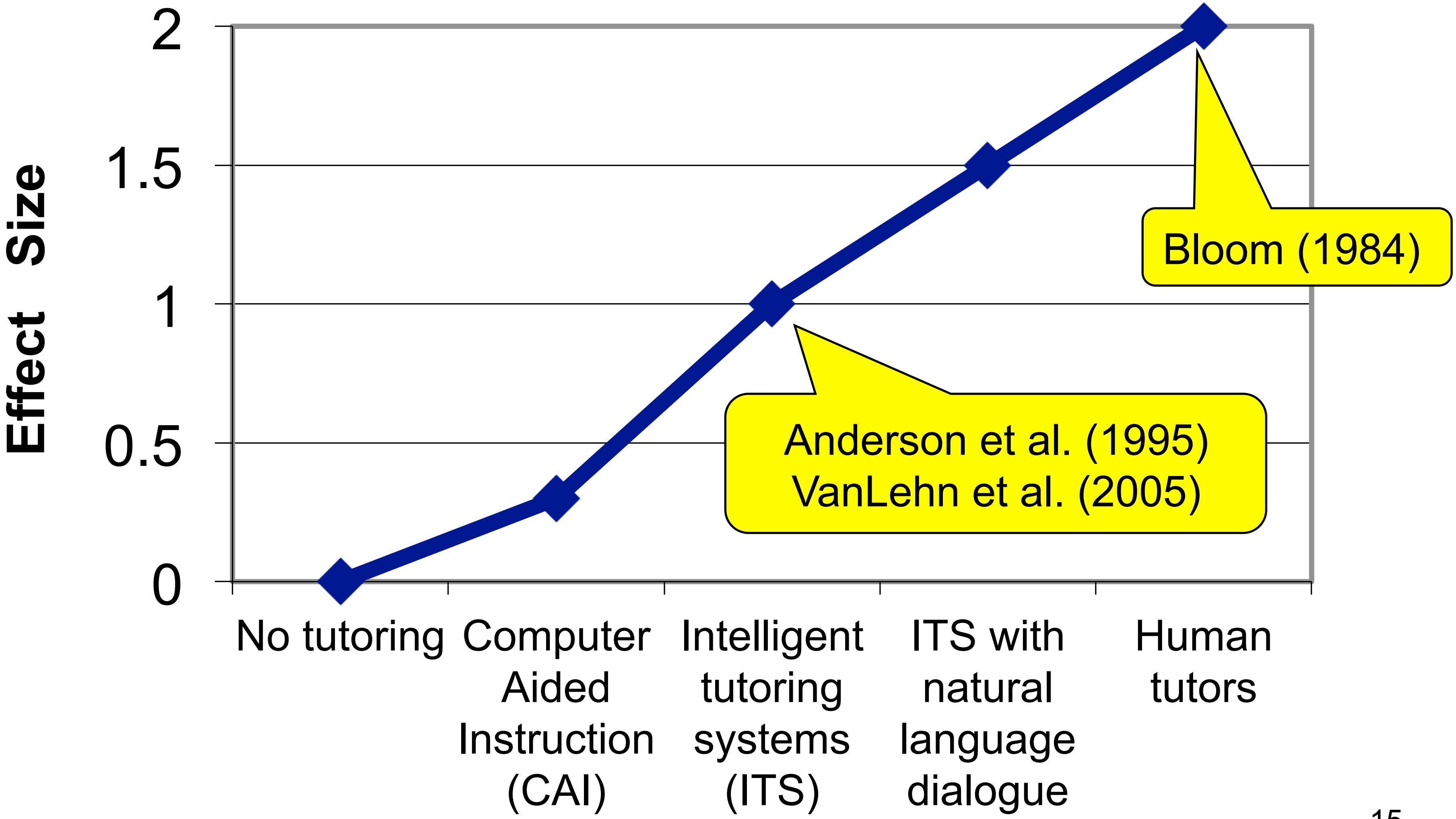
- Types of tutoring systems
- Step-based tutoring ≈ human tutoring
- How to build a step-based tutor
- Increasing their effectiveness
- Flame

A widely held belief: Human tutors are much more effective than computer tutors



$$\frac{\text{Gain(tutored) – Gain(no\_tutor)}}{\text{Standard\_deviation}}$$

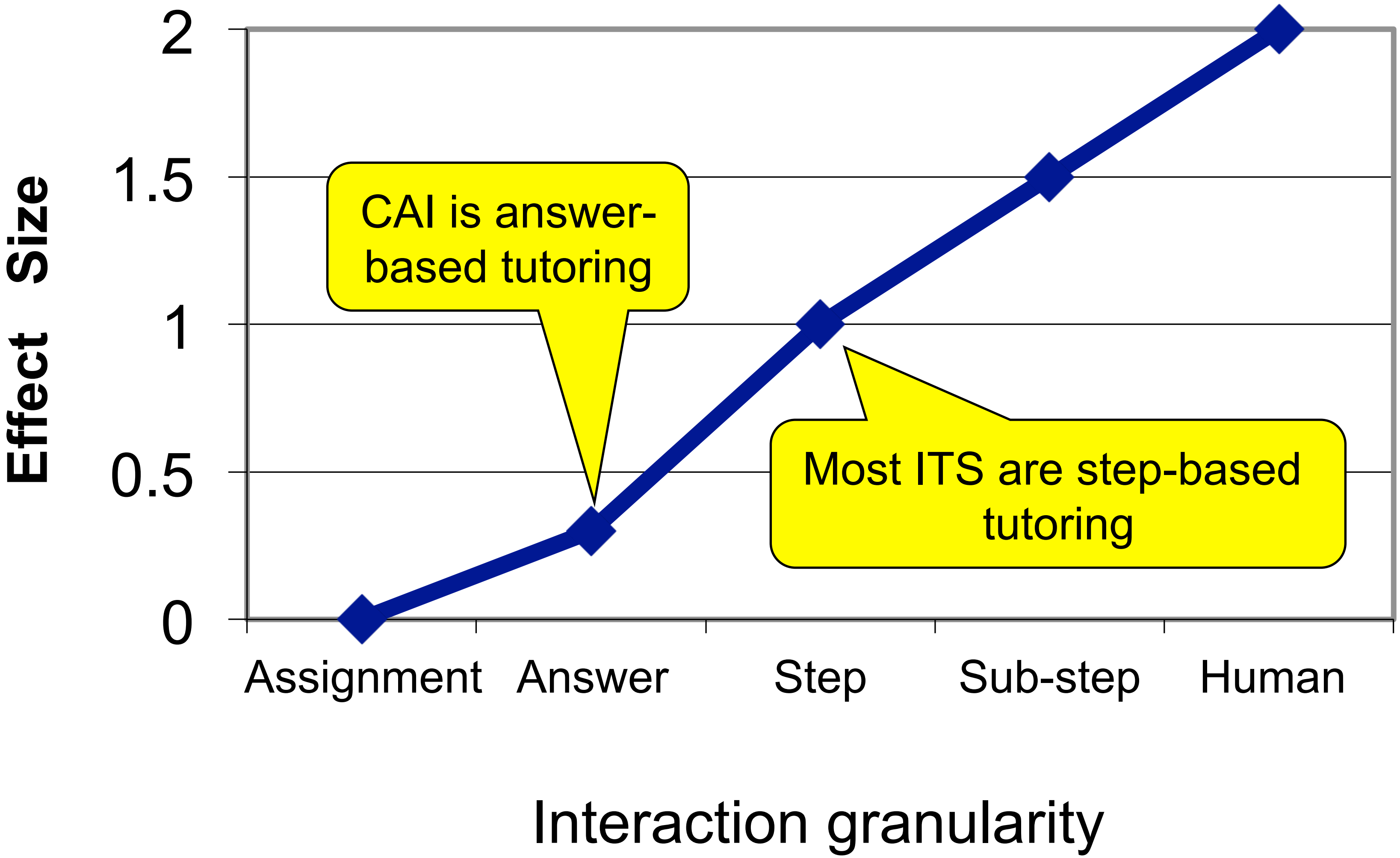
A widely held belief: Human tutors are much more effective than computer tutors



Bloom (1984)

Anderson et al. (1995)
VanLehn et al. (2005)

Common belief: The finer the granularity, the more effective the tutoring

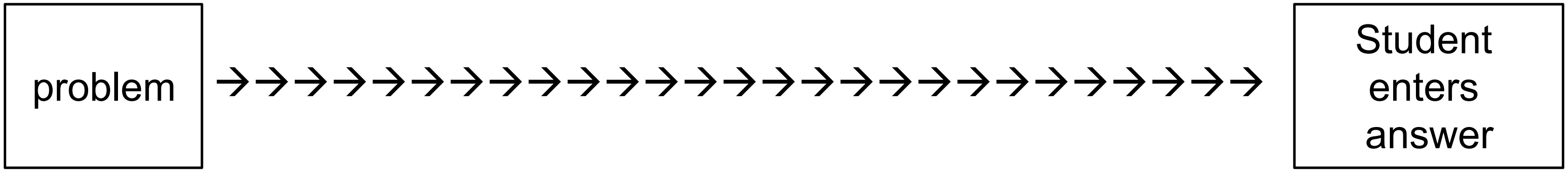


CAI is answer-based tutoring

Most ITS are step-based tutoring

Granularity of tutoring ≈ number of inferences (→) between interactions

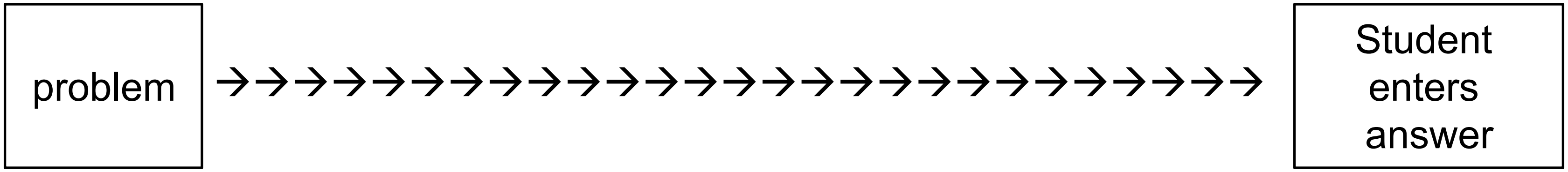
- Answer-based tutoring (CAI)



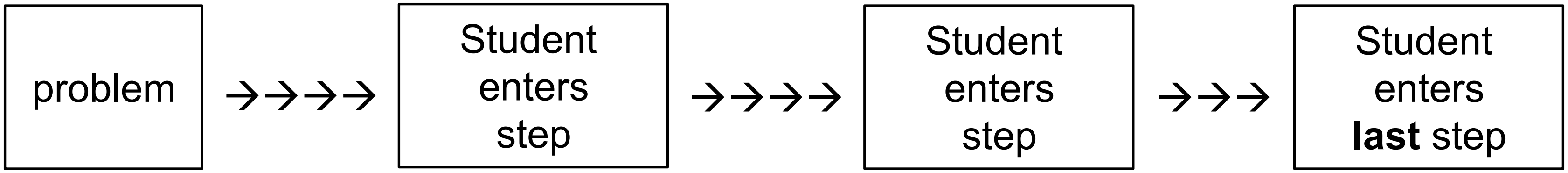
problem →→→→→→→→→→→→→→→→→→→→→→→→→→→ Student enters answer

Granularity of tutorial interaction ≈ number of inferences (→) between interactions

- Answer-based tutoring (CAI)

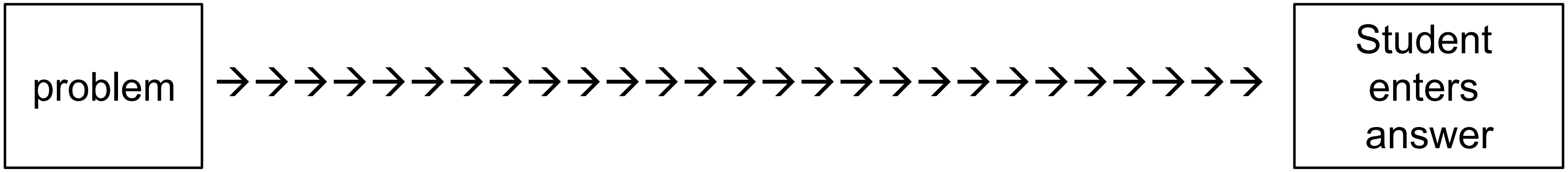


- Step-based tutoring (ITS with ordinary GUI)



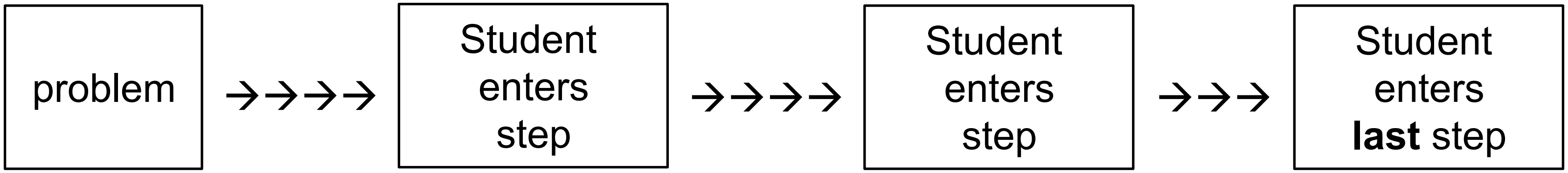
Granularity of tutorial interaction ≈ number of inferences (→) between interactions

- Answer-based tutoring (CAI)



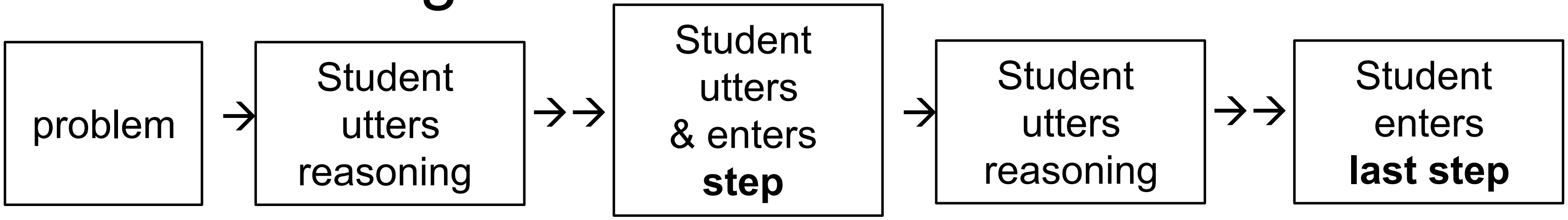
problem →→→→→→→→→→→→→→→→→→→→→→→→→→→ Student enters answer

- Step-based tutoring (ITS with ordinary GUI)



problem →→→→ Student enters step →→→→ Student enters step →→→ Student enters **last** step

- Human tutoring



problem → Student utters reasoning →→ Student utters & enters **step** → Student utters reasoning →→ Student enters **last step**

Hypothesis: The smaller the grain size of interaction, the more effective the tutoring

- Because negative feedback is more effective
  - The shorter the chain of inferences,
    the easier to find the mistake in it
- Because hinting and prompting are more effective
  - The shorter the chain of inferences,
    the easier to infer them from a hint or prompt

Evidence for an interaction **plateau**

- 2 studies from my lab
- 3 studies from other labs
- A meta-analysis

Dialogue & text have same content

Dialogue of Andes-Atlas

T: Here are a few things to keep in mind when computing the acceleration vector for a body at rest. Acceleration is change in *what* over time?

S: velocity

T: Right. If the velocity is *not* changing, what is the magnitude of the acceleration?

S: zero

T: Sounds good. ....

Text of Andes

Here are a few things to keep in mind when calculating acceleration for a body at rest. Acceleration is change in velocity over time. If velocity is not changing, then there is zero acceleration....

Results comparing Andes-Atlas to Andes

- Study 1: Andes-Atlas > Andes
  - but content not controlled properly
- Study 2 (N=26): Andes-Atlas ≈ Andes (p>.10)
- Study 3 (N=21): Andes-Atlas < Andes (p<.10, d=0.34)
- Study 4 (N=12): Andes-Atlas ≈ Andes (p>.10)

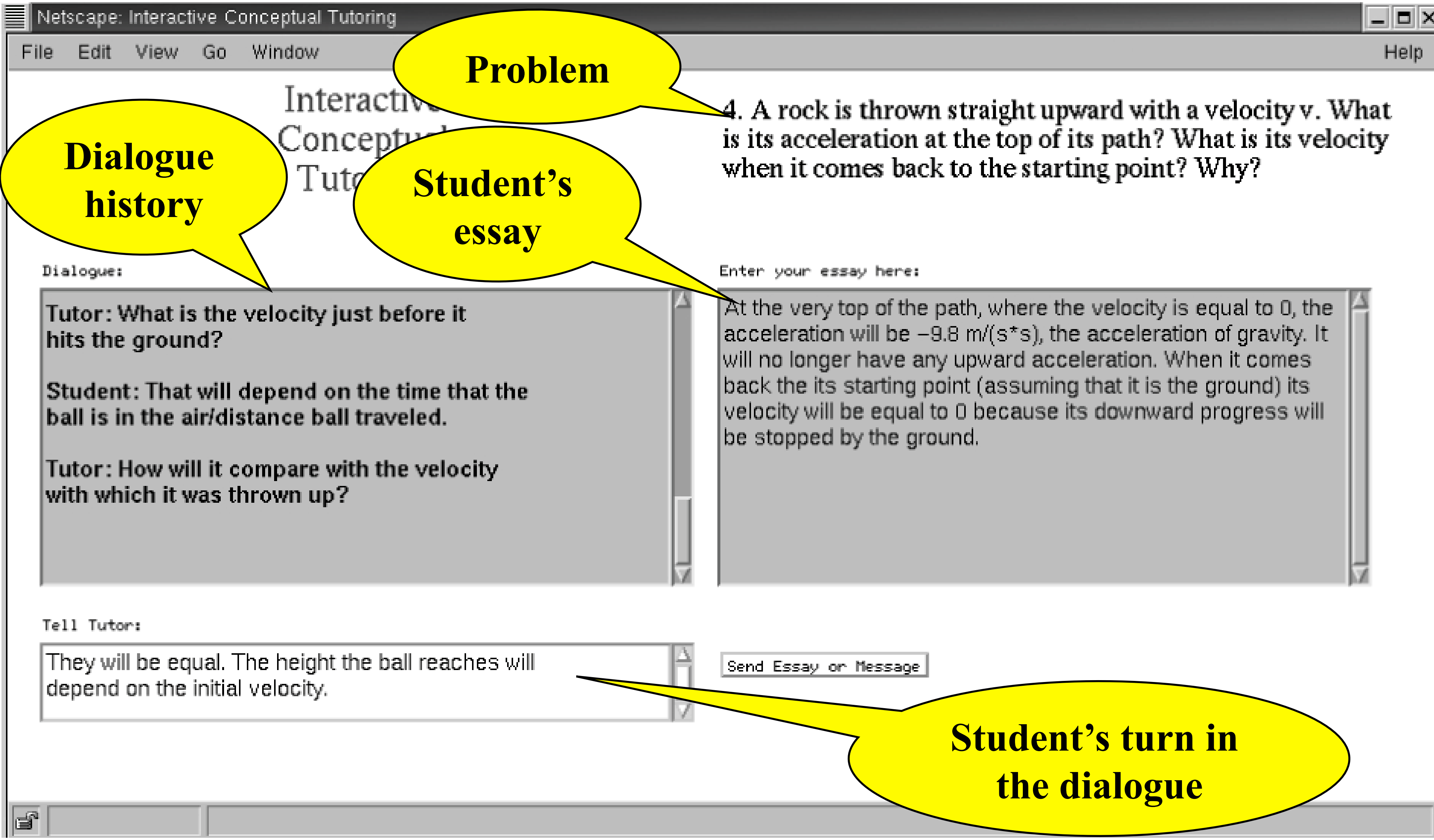
Conclusion: Substep tutoring is *not* more effective than step-based tutoring

The WHY2 studies

(VanLehn, Graesser et al., 2007, *Cognitive Science*)

- 5 conditions
  - Human tutors
  - Substep-based tutoring system
    - Why2-Atlas
    - Why2-AutoTutor (Graesser et al.)
  - Step-based tutoring system
  - Text
- Procedure
  - Pretraining
  - Pre-test
  - Training (~ 4 to 8 hours)
  - Post-test

User interface for human tutoring and Why2-Atlas



Problem

4. A rock is thrown straight upward with a velocity v. What is its acceleration at the top of its path? What is its velocity when it comes back to the starting point? Why?

Dialogue history

Student's essay

Dialogue:

Tutor: What is the velocity just before it hits the ground?

Student: That will depend on the time that the ball is in the air/distance ball traveled.

Tutor: How will it compare with the velocity with which it was thrown up?

Enter your essay here:

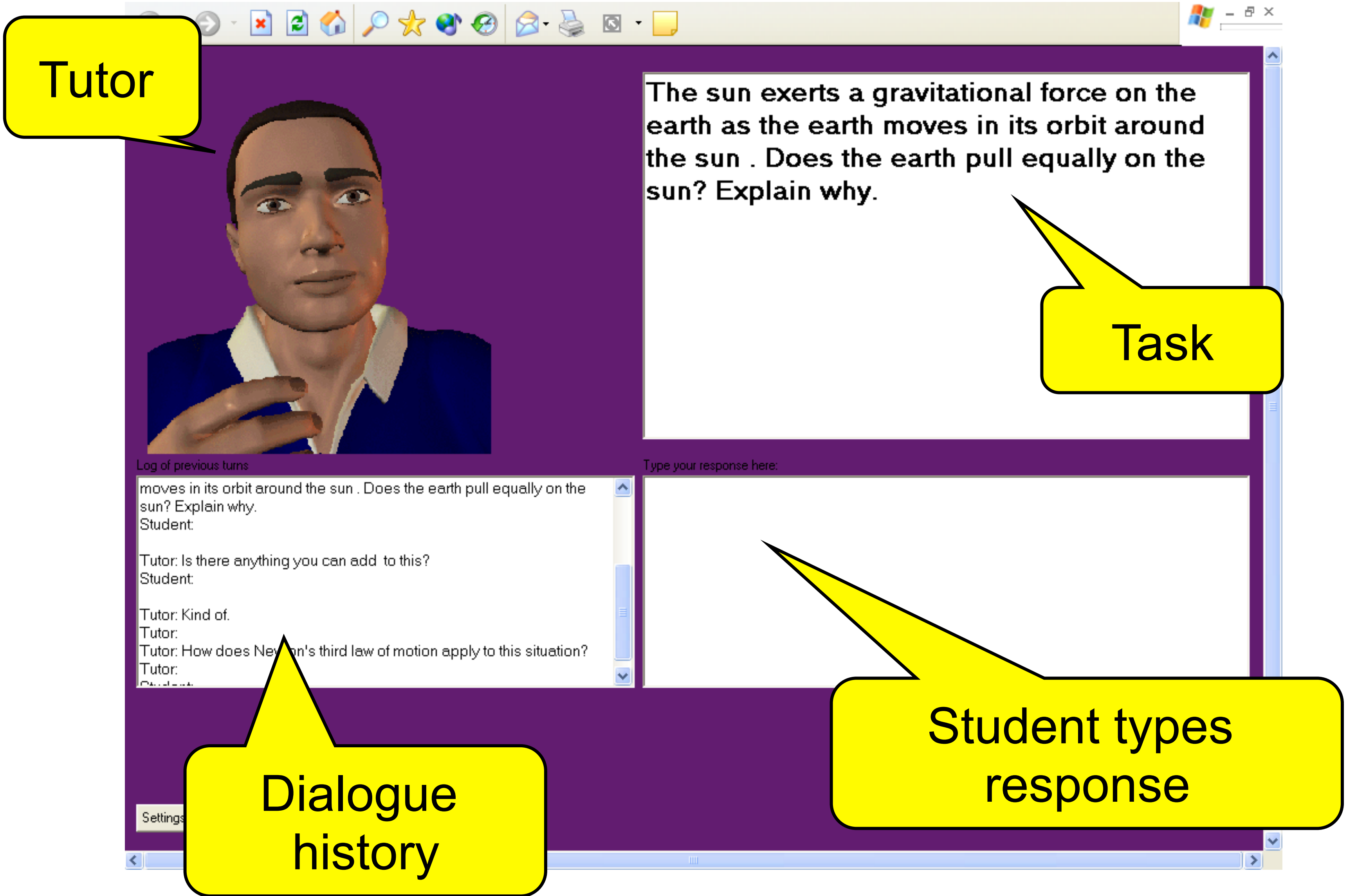
At the very top of the path, where the velocity is equal to 0, the acceleration will be –9.8 m/(s*s), the acceleration of gravity. It will no longer have any upward acceleration. When it comes back the its starting point (assuming that it is the ground) its velocity will be equal to 0 because its downward progress will be stopped by the ground.

Tell Tutor:

They will be equal. The height the ball reaches will depend on the initial velocity.

Student's turn in the dialogue

Why2-AutoTutor user interface



The sun exerts a gravitational force on the earth as the earth moves in its orbit around the sun . Does the earth pull equally on the sun? Explain why.

Log of previous turns

moves in its orbit around the sun . Does the earth pull equally on the sun? Explain why.
Student:

Tutor: Is there anything you can add to this?
Student:

Tutor: Kind of.
Tutor:
Tutor: How does Newton's third law of motion apply to this situation?
Tutor:
Student:

Type your response here:

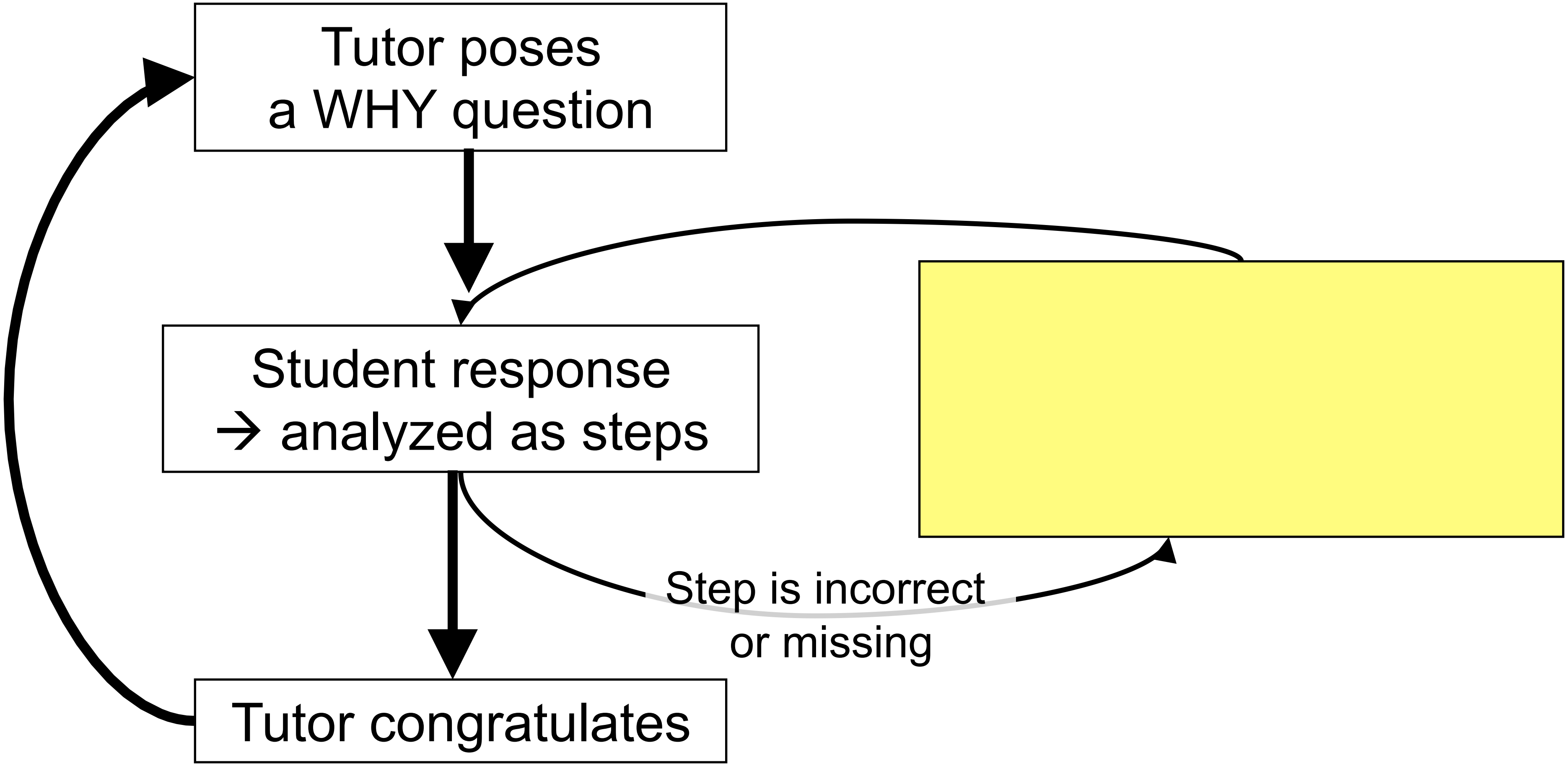
Tutor

Task

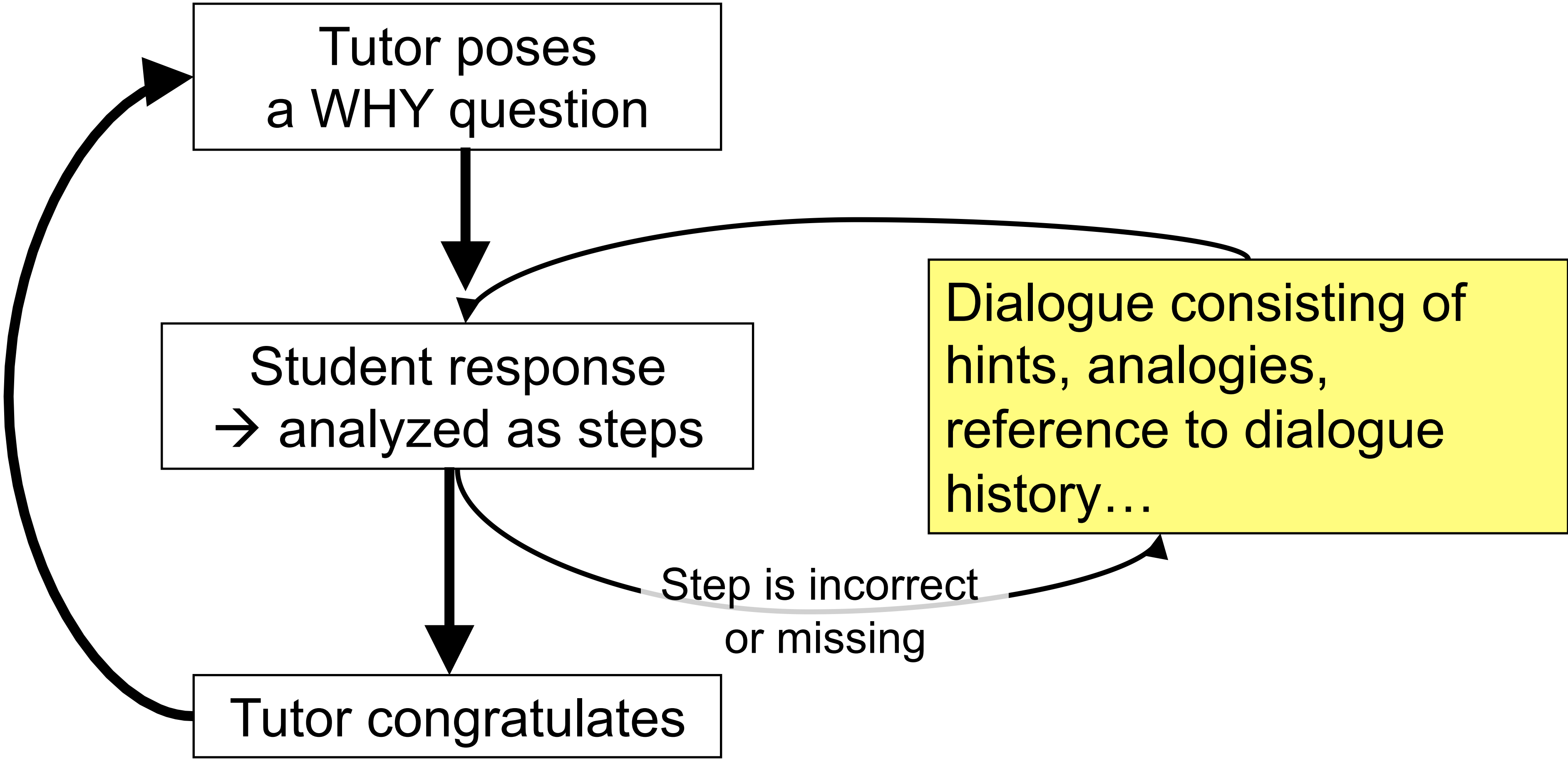
Dialogue history

Student types response

Only difference between tutoring conditions was contents of yellow box


Tutor poses a WHY question
Student response → analyzed as steps
Step is incorrect or missing
Tutor congratulates

Human tutoring



Tutor poses a WHY question

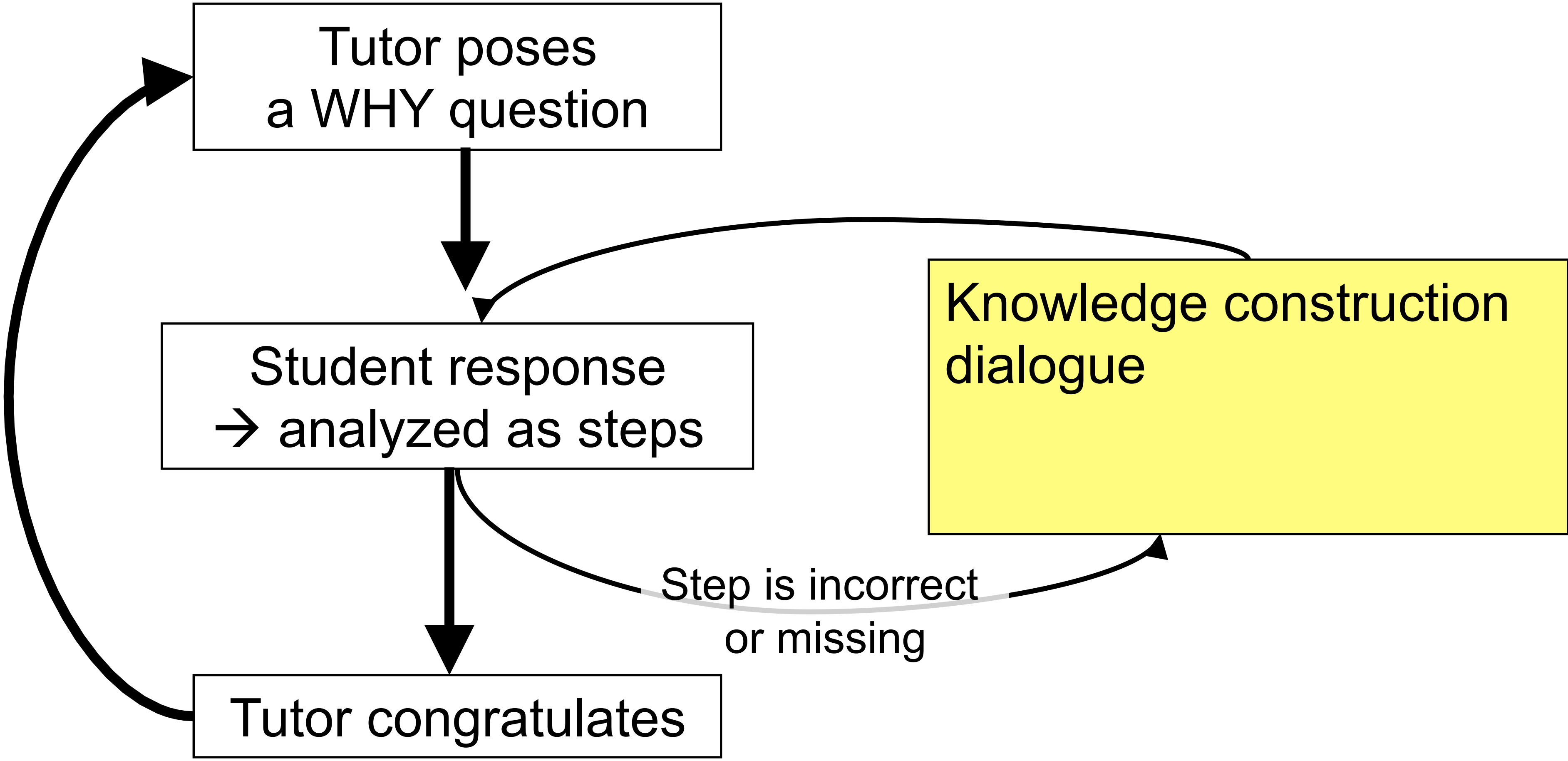
Student response → analyzed as steps

Dialogue consisting of hints, analogies, reference to dialogue history…

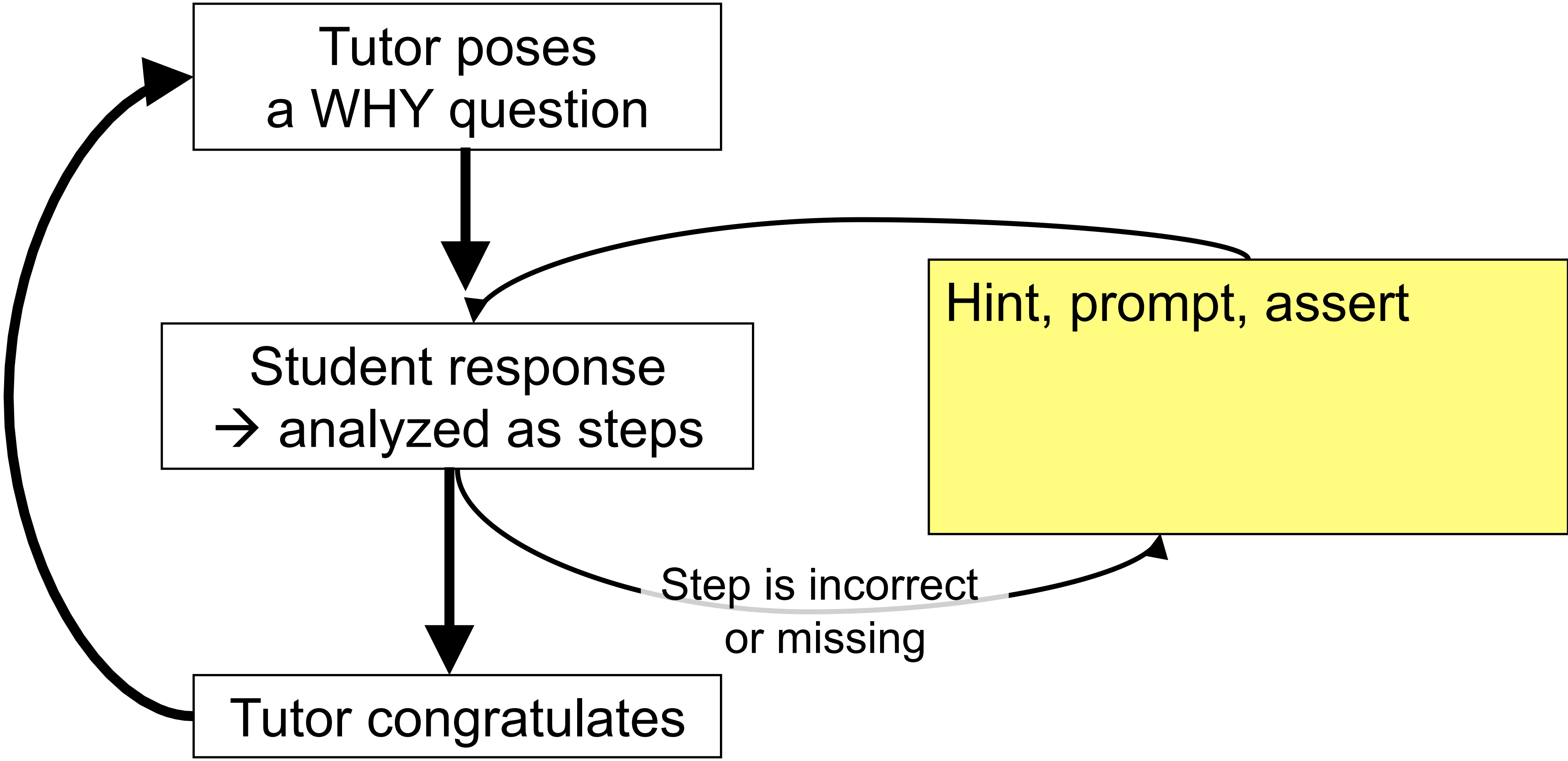
Step is incorrect or missing

Tutor congratulates

Why2-Atlas


Tutor poses a WHY question
Student response → analyzed as steps
Knowledge construction dialogue
Step is incorrect or missing
Tutor congratulates

Why2-AutoTutor



Tutor poses a WHY question

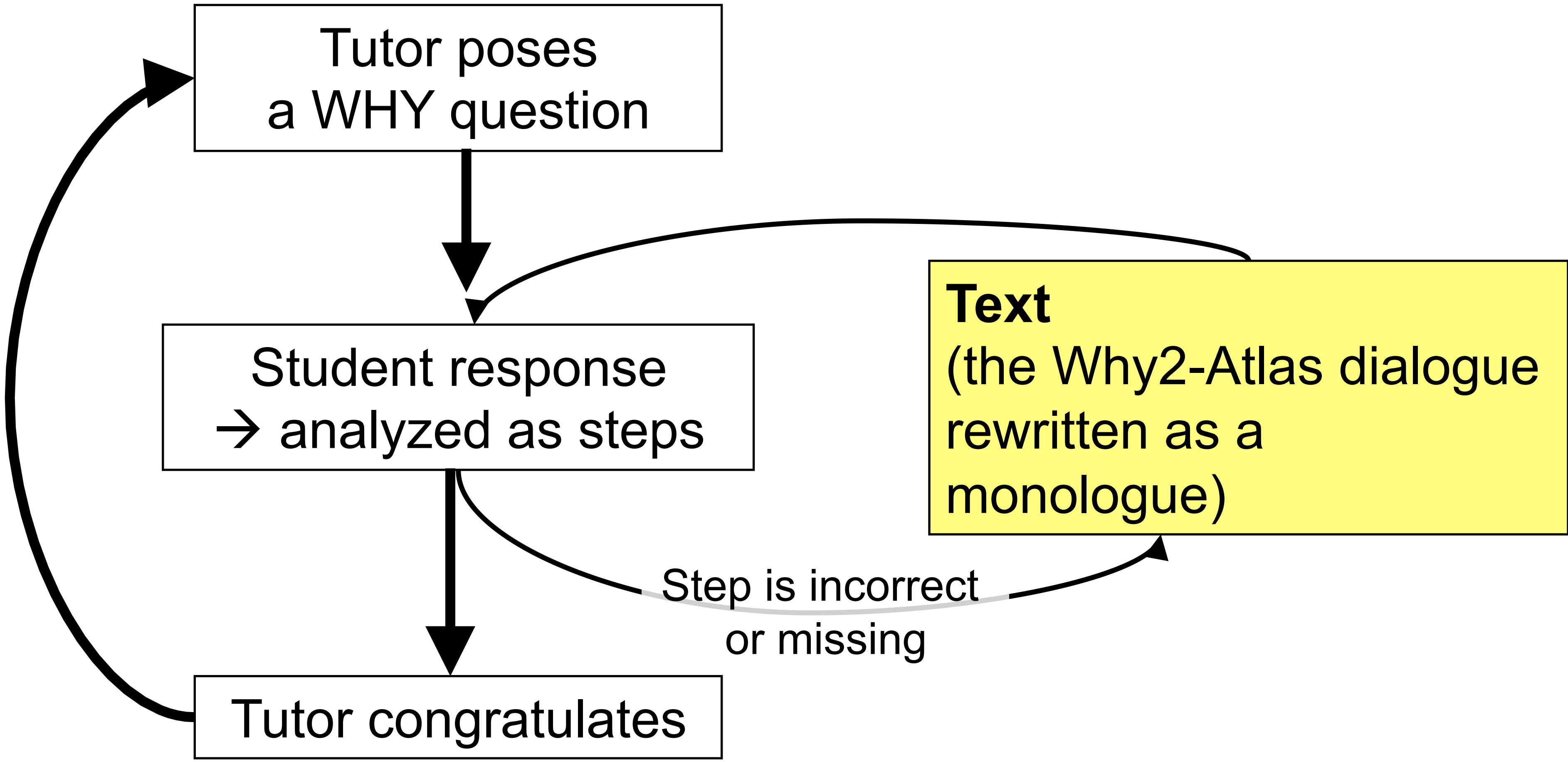
Student response → analyzed as steps

Hint, prompt, assert

Step is incorrect or missing

Tutor congratulates

A step-based tutor: A text explanation with same content



Tutor poses a WHY question

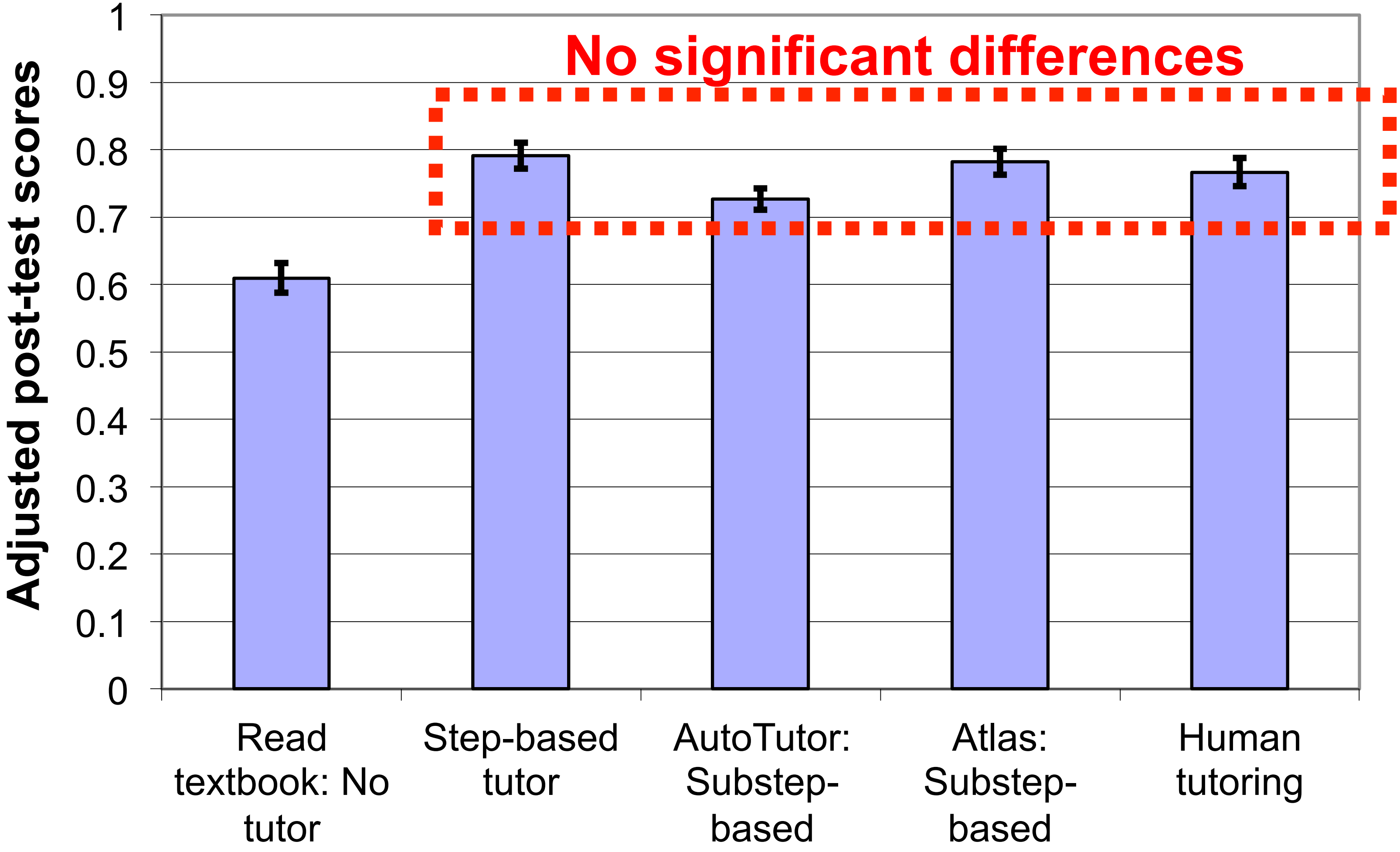
Student response → analyzed as steps

Text
(the Why2-Atlas dialogue rewritten as a monologue)

Step is incorrect or missing

Tutor congratulates

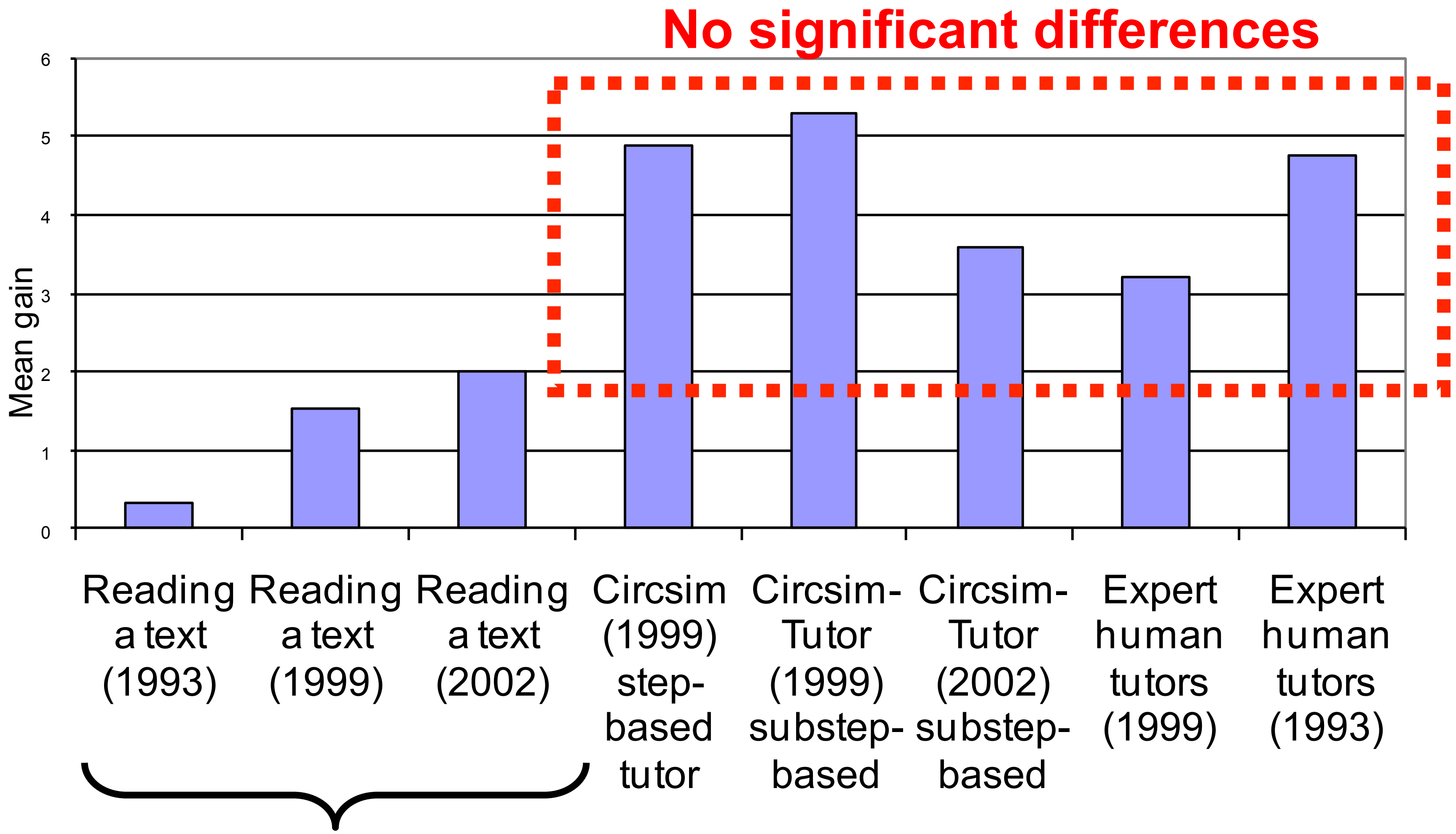
Experiments 1 & 2



Results from all 7 experiments

- Human tutoring
  = Substep-based tutoring systems
  = Step-based tutoring system
- Tutors > Textbook (no tutoring)
- Atlas (symbolic NLP) = AutoTutor (statistical NLP)

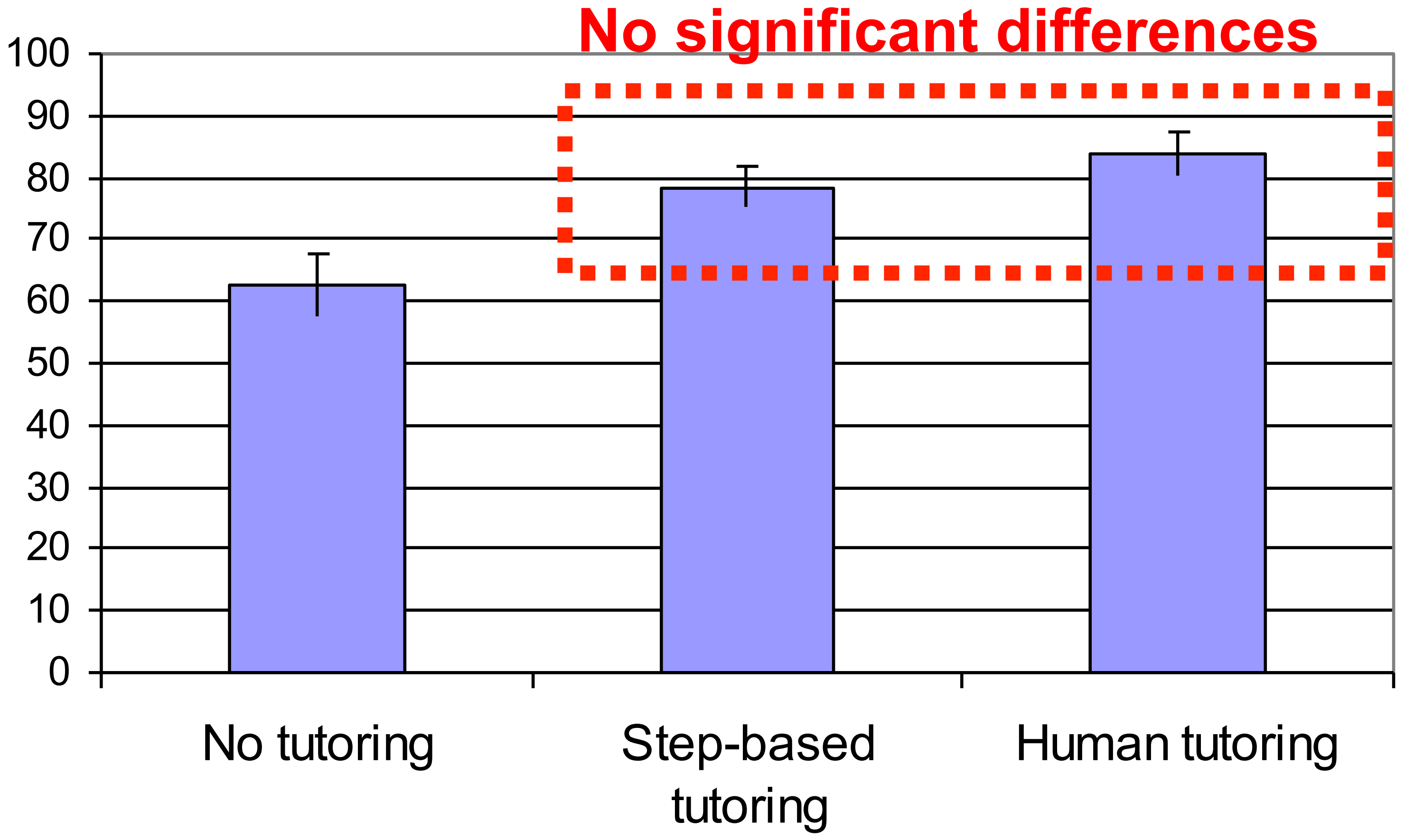
Evens & Michael (2006) also show human tutoring = sub-step-based tutoring = step-based tutoring



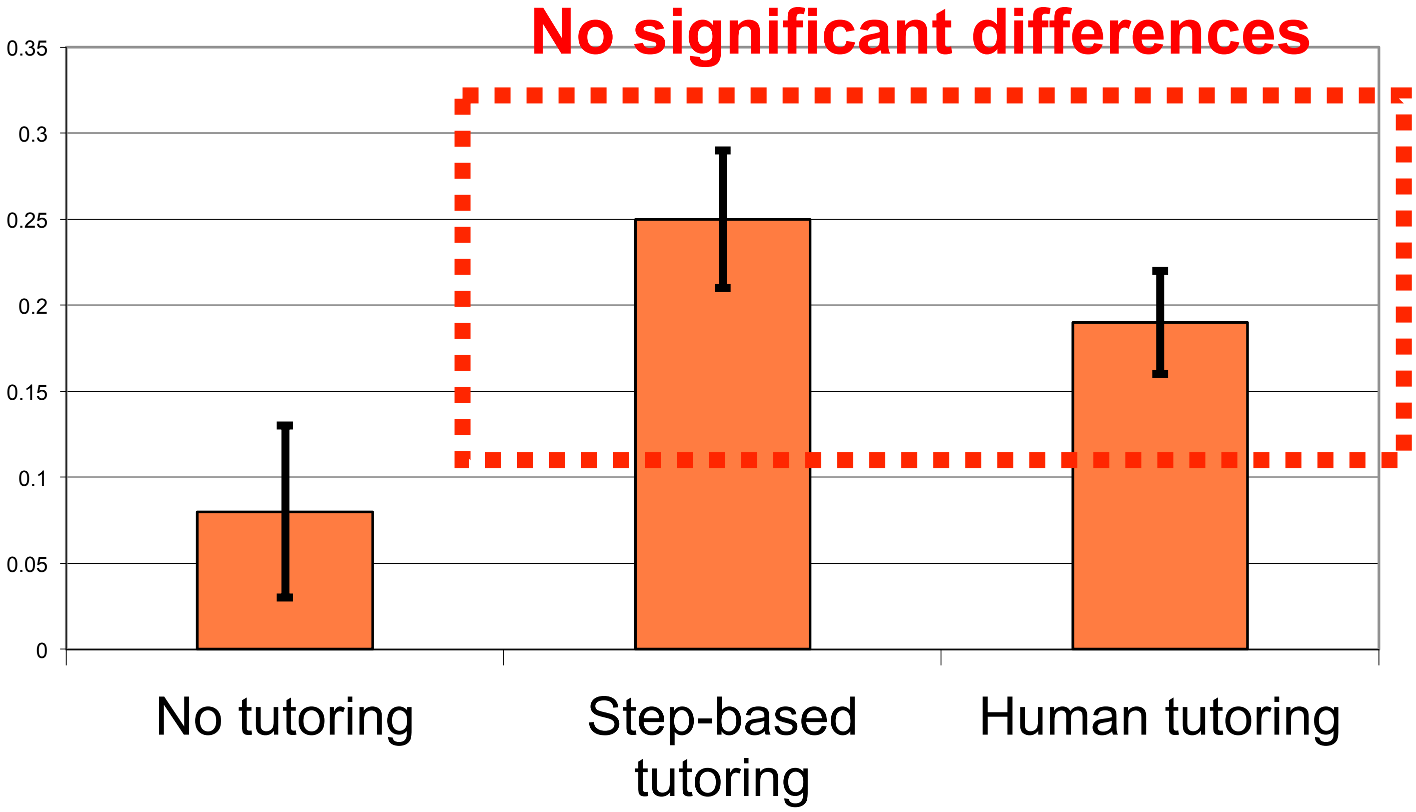
No significant differences

No tutoring

Reif & Scott (1999) also show human tutors = step-based tutoring



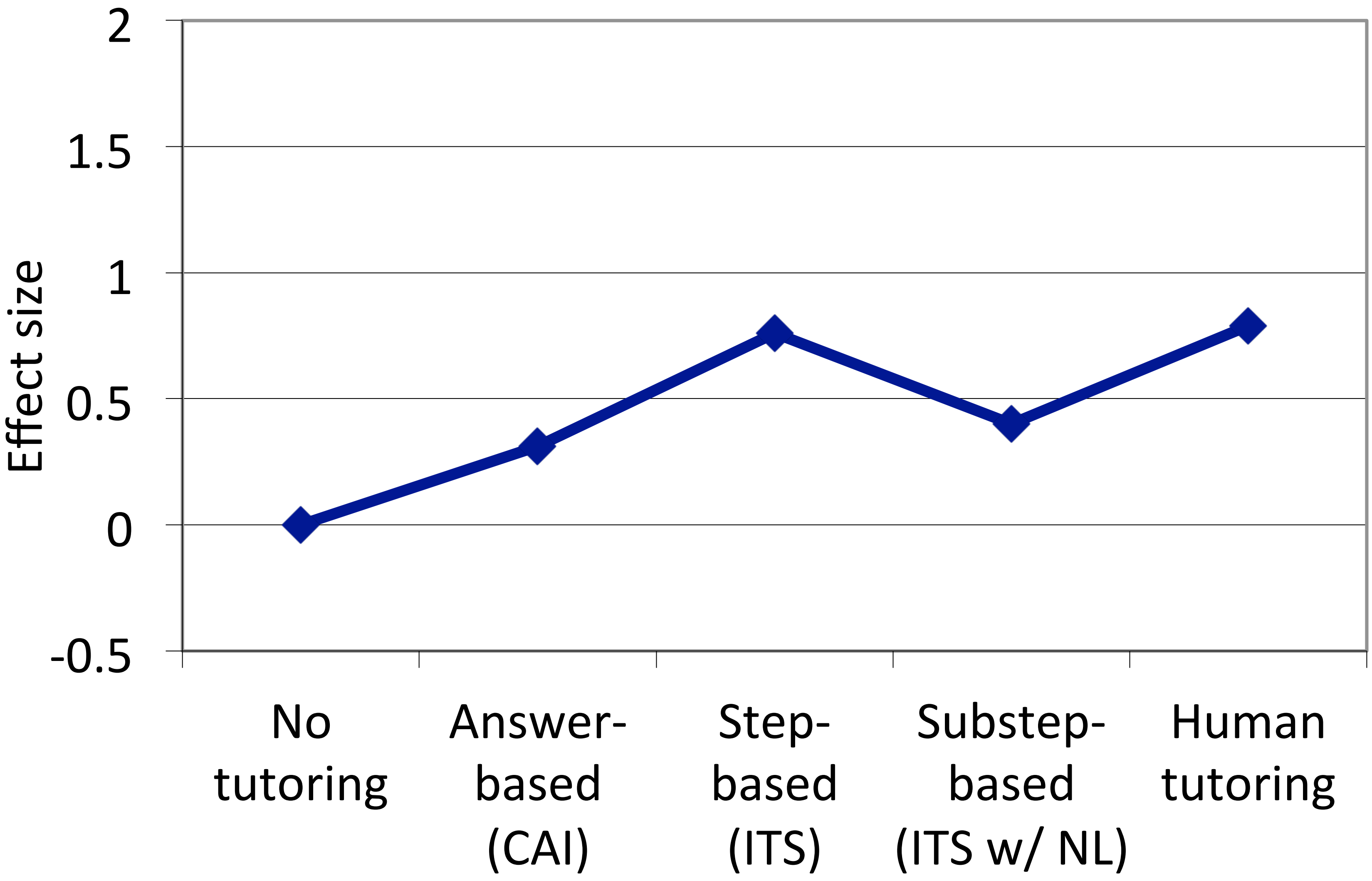
Katz, Connelly & Allbritton (2003) post-practice reflection: human tutoring = step-based tutoring



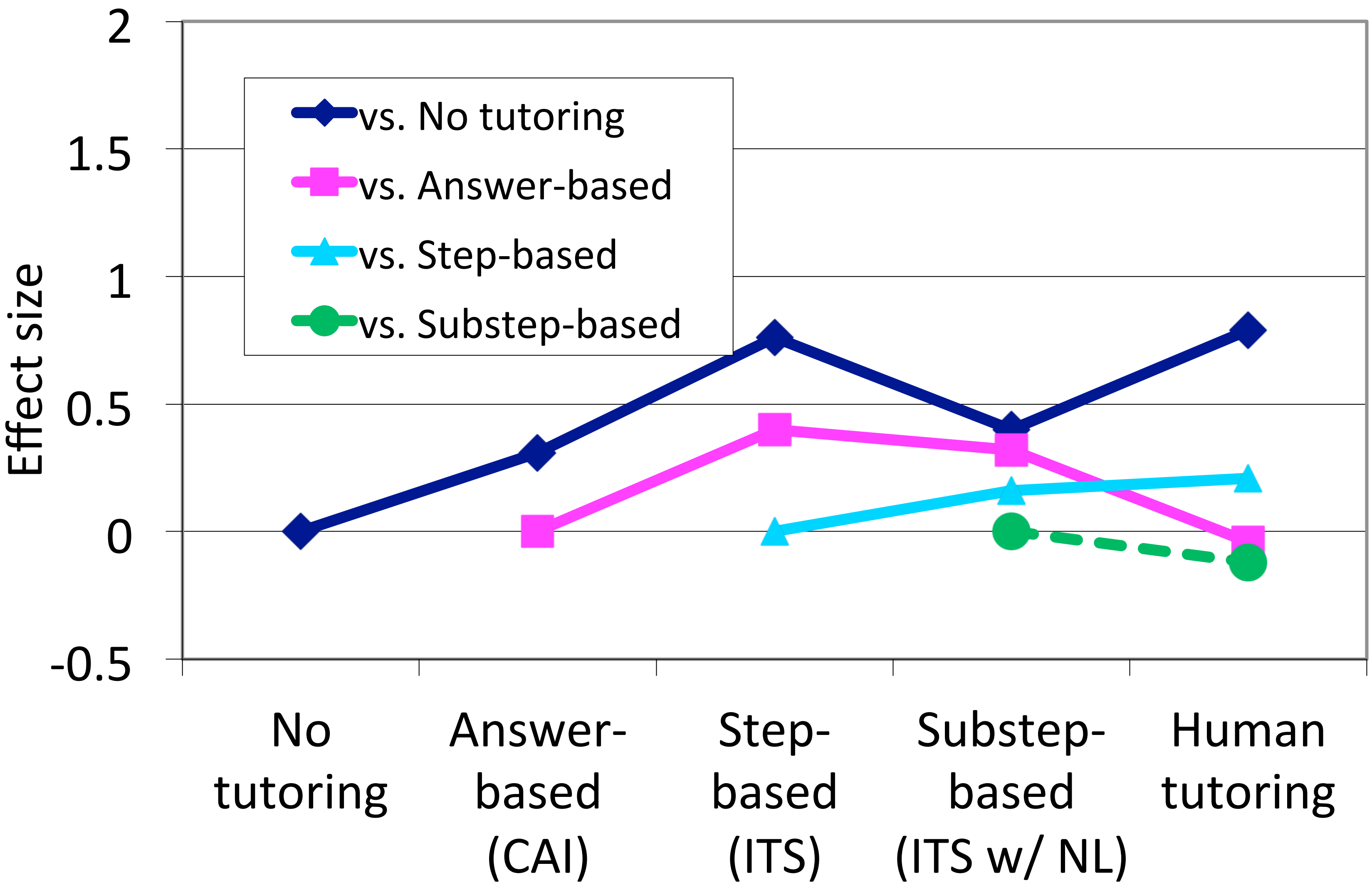
Meta-analytic results for all possible pairwise comparisons (VanLehn, 2011)

Tutoring type	vs. other tutoring type	Num. of effects	Mean effect	% reliable
Answer-based	no tutoring	165	0.31	40%
Step-based		28	0.76	68%
Substep-based		26	0.40	54%
Human		10	0.79	80%
Step-based	answer-based	2	0.40	50%
Substep-based		6	0.32	33%
Human		1	-0.04	0%
Substep-based	step-based	11	0.16	0%
Human		10	0.21	30%
Human	sub-step based	5	-0.12	0%

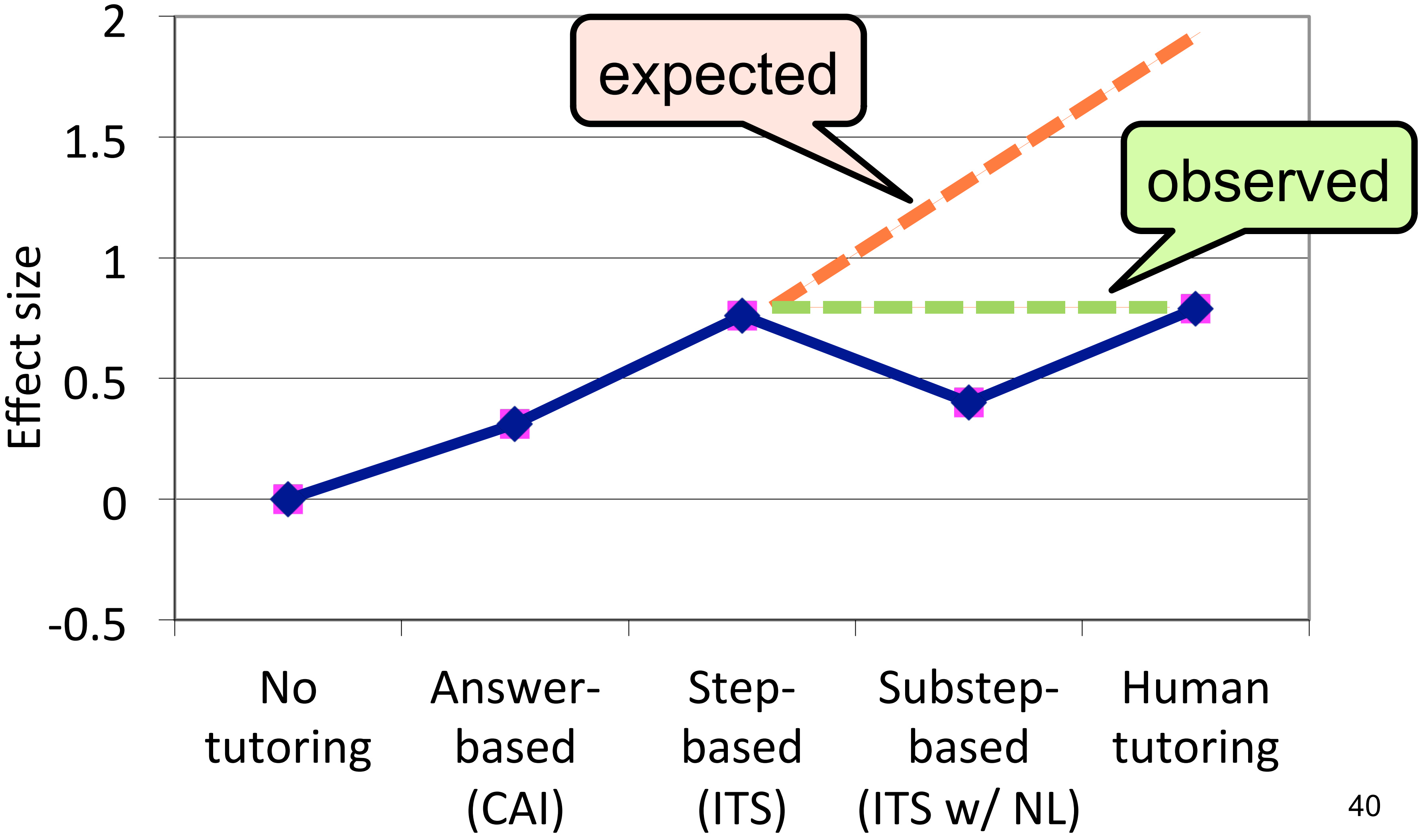
Graph of comparisons of 4 tutoring types vs. no tutoring



Graphing all 10 comparisons: No tutor < CAI < ITS = ITS w/NL = human



Graph of comparisons of 4 tutoring types vs. no tutoring



The interaction **plateau** hypothesis

- The smaller the grain size of interaction, the more effective the tutoring
  - Assignments < answers < steps
- But grain sizes less than steps are no more effective than steps
  - Steps = substeps = human

Limitations & caveats

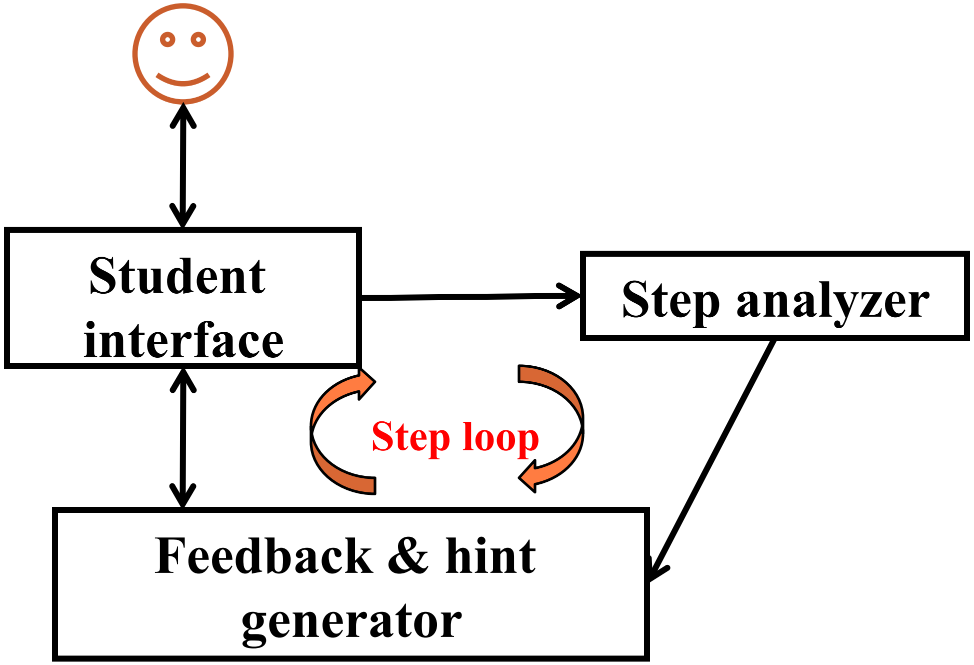
- Task domain
  - Must allow computer tutoring
  - Only STEM; not language, music, sports...
- Normal learners
  - Not learning disabled
  - Prerequisite knowledge mastered
- Human tutors must teach **same content** as computer tutors
  - Only the type of tutoring (human, ITS, CAI) varies
- One-on-one tutoring

Outline

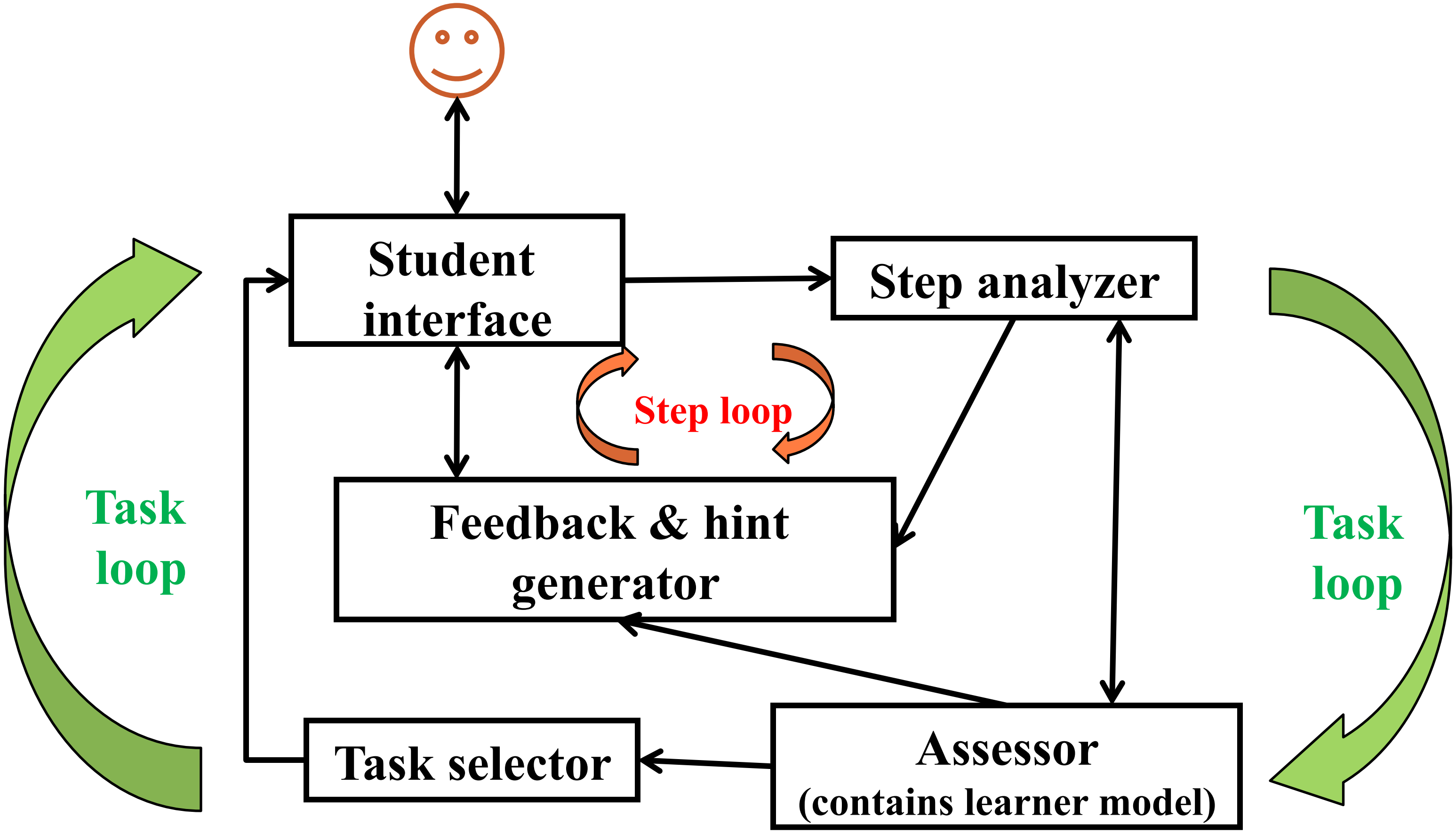
Next

- Types of tutoring systems
- Step-based tutoring ≈ human tutoring
- How to build a step-based tutor
- Increasing their effectiveness
- Flame

Main modules of a non-adaptive step-based tutoring system



Main modules of an adaptive step-based tutoring system



Main types of step analyzers

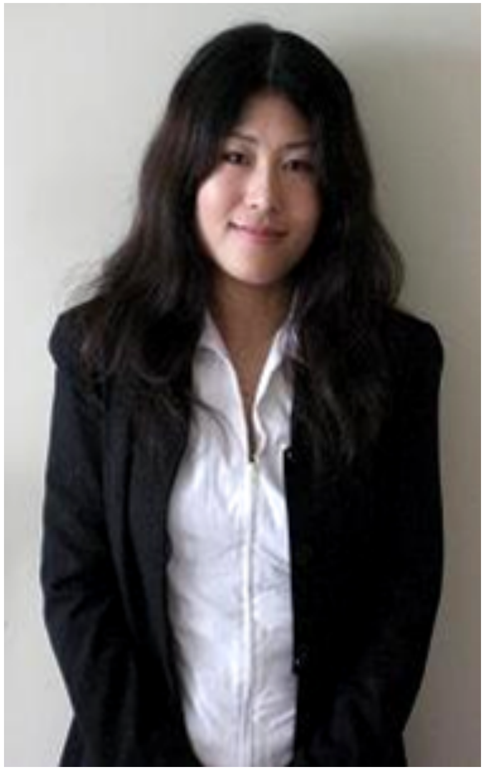
- Three main methods for generating ideal steps
  - *Model tracing*: One expert system that can solve all problems in all ways
  - *Example tracing*: For each problem, all acceptable solutions
  - *Constraint-based:* Example + recognizers of bad steps + recognizers of steps equivalent to example's steps
- Comparing student and ideal steps
  - Trivial if steps are menu choices, numbers, short texts
  - Harder if steps are math, logic, chemistry, programming
  - Use statistical NLP for essays, long explanations
  - Use probabilistic everything for gestures

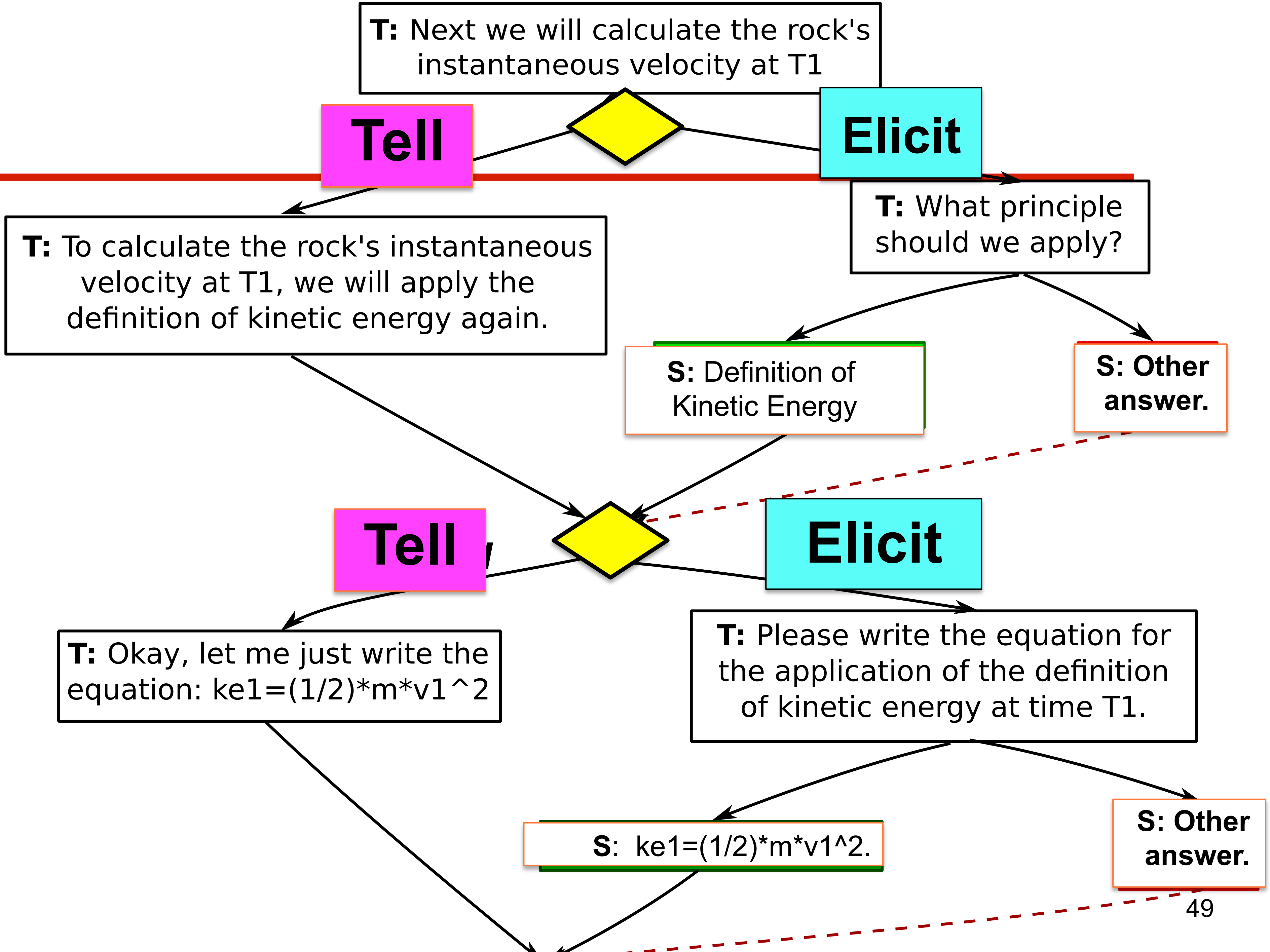
Outline

- Types of tutoring systems
- Step-based tutoring ≈ human tutoring
- How to build a step-based tutor
- Next → Increasing their effectiveness
- Flame

The details can make a huge difference. How can we get them right?

- Called A/B testing in the game industry
- During example-based tutoring, when should the tutor ***tell*** the student an inference vs. ***elicit*** it from the student?
- Can machine-learned policies improve the ***tell vs. elicit*** decision?
- Min Chi's Ph. D. thesis





T: Next we will calculate the rock's instantaneous velocity at T1

Tell

Elicit

T: To calculate the rock's instantaneous velocity at T1, we will apply the definition of kinetic energy again.

T: What principle should we apply?

S: Definition of Kinetic Energy

S: Other answer.

Tell

Elicit

T: Okay, let me just write the equation: ke1=(1/2)*m*v1^2

T: Please write the equation for the application of the definition of kinetic energy at time T1.

S: ke1=(1/2)*m*v1^2.

S: Other answer.

5-Stage Procedure

Stage 1	Study: 64 students using random policy.
Stage 2	Calculate Sub-optimal policy.
Stage 3	Study: 37 students using Sub-optimal policy
Stage 4	Calculate Enhancing & Diminishing policies.
Stage 5	Study: 29 students using Enhancing policy vs. 28 students using Diminishing policy

Diminishing policy is calculated to *decrease* learning.
Other policies are calculated to *increase* learning.

Calculated policies are composed of many rules, such as:

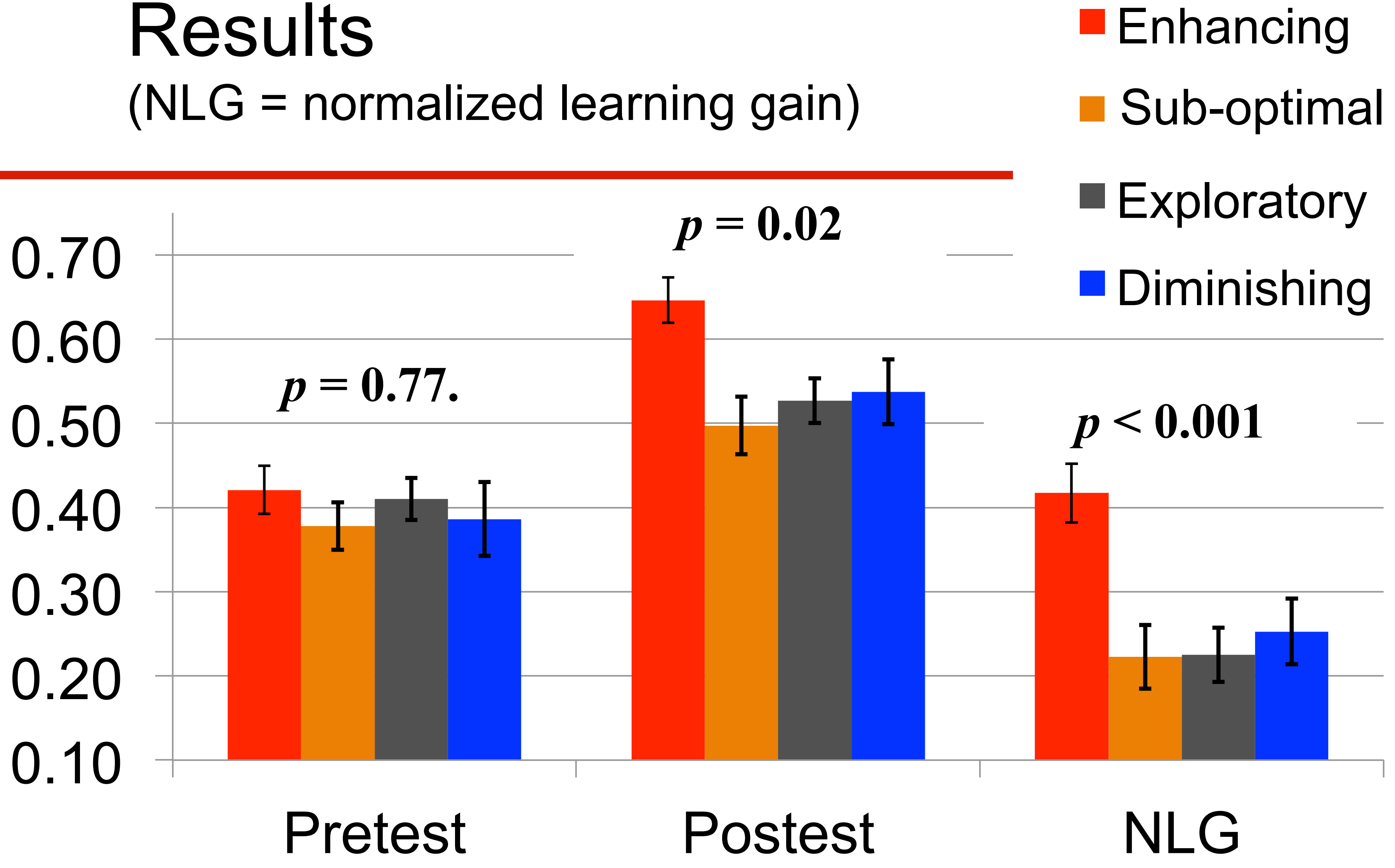
If problem: difficult
And last tutor action: tell
And student performance: high
And duration since last mention of the current principle ≥ 50 sec

→ **Elicit**

Machine learner selected features in left side of rule from 50 possible features defined by humans

Results

(NLG = normalized learning gain)



Enhancing > everything else, which were about the same

Conclusions’ from Min Chi’s thesis

- Details do matter e.g., the Tell vs. Elicit decision
- Improved policies for Tell vs. Elicit can be induced from *modest amounts of data*
  - 103 students
- Induced policies can have a large effect on learning gains (d=0.8).
- Developers should do many such A/B studies

Overall conclusion: We need to use more step-based tutors

	Non-adaptive	**Competency gating**	**Adaptive task selection**
Answer-based feedback/hints	Thousands	Hundreds	Few
Step-based feedback/hints	Hundreds (few on market)	Tens	Few
Sub-step based feedback/hints	Tens	None	None

Outline

- Types of tutoring systems
- Step-based tutoring ≈ human tutoring
- How to build a step-based tutor
- Increasing their effectiveness
- Flame

Next

Why are there so few step-based tutoring systems?

- K-12 curriculum and standardized tests have evolved to favor answer-based tasks
- K-12 instructors do not view homework as the problem area; it's classroom time that concerns them.
- Instructors need to share knowledge, policies and authority with a tutoring system

Why are competency-gated tutoring systems so rare?

- Schools are time-gated, not competency-gated
- Difficulty enforcing deadlines
- Grading based on time-to-mastery may be pointless and harmful.

Recommendation for instructors

- Use competency-gated tutoring system
  - Flip: Videos/reading at home. Exercises in class.
  - Half group work (paper?) and half individual work (tutor)
  - Noisy study halls instead of lecture halls
  - Deadlines & exams for core. Badges for enrichment.
- Use a step-based tutoring system
  - Buy one if you can
  - If you build one, use example-tracing first
  - If you will use it repeatedly, plan on A/B testing

Recommendations for parents

- Human tutors ≈ step-based tutoring systems
- If you can do the task, then you can tutor the task
  - Do not lecture/demo!
  - Be step-based.

Thank you!

Bibliography
(all papers available from public.asu.edu/~kvanlehn)

- The meta-analysis
  - VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist, 46(4), 197-221.*
- Why2 experiments
  - VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31(1), 3-62.*
- Andes, the physics tutor
  - VanLehn, K., Lynch, C., Schultz, K., Shapiro, J. A., Shelby, R. H., Taylor, L., et al. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education, 15(3), 147-204.*
- Andes-Cordillera study
  - in prep

Bibliography continued

- Andes-Atlas studies
  - Siler, S., Rose, C. P., Frost, T., VanLehn, K., & Koehler, P. (2002, June). *Evaluating knowledge construction dialogues (KCDs) versus minilesson within Andes2 and alone. Paper presented at the Workshop on dialogue-based tutoring at ITS 2002, Biaritz, France.*
- Machine learning of Cordillera policies
  - Chi, M., VanLehn, K., Litman, D., & Jordan, P. (2011). Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. *User Modeling and User-Adapted Interaction, 21(1-2), 99-135.*
- *Teaching a meta-cognitive strategy (MEA)*
  - Chi, M., & VanLehn, K. (2010). Meta-cognitive strategy instruction in intelligent tutoring systems: How, when and why. *Journal of Educational Technology and Society, 13(1), 25-39.*