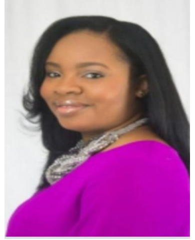
An Astonishing Regularity in Student Learning Rate and the Benefits of Formative Assessment …

Ken Koedinger with great help from many co-authors …



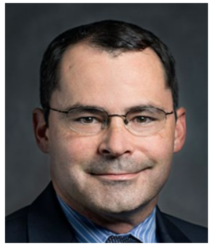




Paulo Carvalho Mimi McLaughlin Norman Bier Ran Liu Cassandra Brentley







Danielle Chine/Thomas Liz Richey Menaf Gul Carmen Thomas-Browne Lee Branstetter

Thanks to **NSF** for funding of LearnLab & LearnSphere and to **Gates, , RK Mellon, Schmidt** for funding of PLUS

Overview

Interactive learning by doing & online tutors

What student differences account for learning outcome differences?

Koedinger, Carvalho, Liu, McLaughlin (2023). An Astonishing Regularity in Student Learning Rate. *PNAS*.

Formative Assessment instead of Summative Assessment

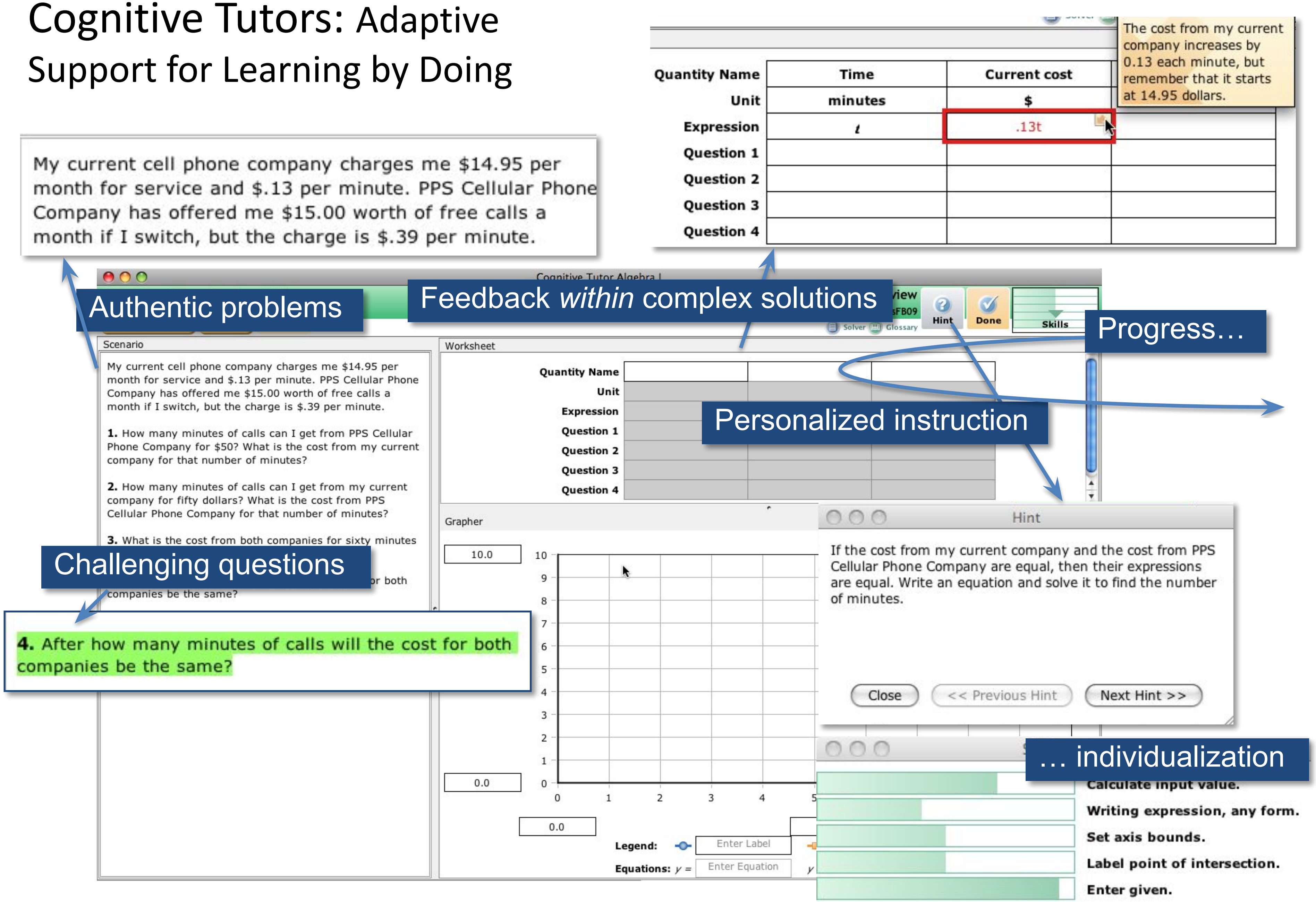
Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*.

Can hybrid human-AI tutoring enhance educational equity?

Chine et al (2022). Educational Equity Through Combined Human-AI Personalization. AIEd Conference.

Holstein et al (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

Cognitive Tutors: Adaptive Support for Learning by Doing



Cognitive Tutors: Adaptation techniques

- Cognitive Model: Computer simulation of student thinking

$$3(2x - 5) = 9$$

If goal is solve a(bx+c) = d
Then rewrite as abx + ac = d

If goal is solve a(bx+c) = d
Then rewrite as bx+c = d/a

If goal is solve a(bx+c) = d
Then rewrite as abx + c = d

$6x - 15 = 9$ $\qquad$ $2x - 5 = 3$ $\qquad$ $6x - 5 = 9$

Two algorithms use *Cognitive Model* to adapt:

- Model Tracing: Follows each student's thinking steps
  => context-sensitive instruction

Koedinger et al. (2013). [New potentials for data-driven intelligent tutoring system development and optimization](). *AI Magazine.*

Cognitive Tutors: Adaptation techniques

- Cognitive Model: Computer simulation of student thinking

$3(2x - 5) = 9$

If goal is solve $a(bx+c) = d$
Then rewrite as $abx + ac = d$

Hint message: "Distribute a across the parentheses."

Known? = 85% chance

$6x - 15 = 9$

Known? = 45%

$2x - 5 = 3$

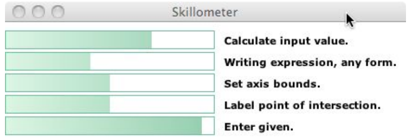
If goal is solve $a(bx+c) = d$
Then rewrite as $abx + c = d$

Bug message: "You need to multiply c by a also."

$6x - 5 = 9$

Two algorithms use *Cognitive Model* to adapt:

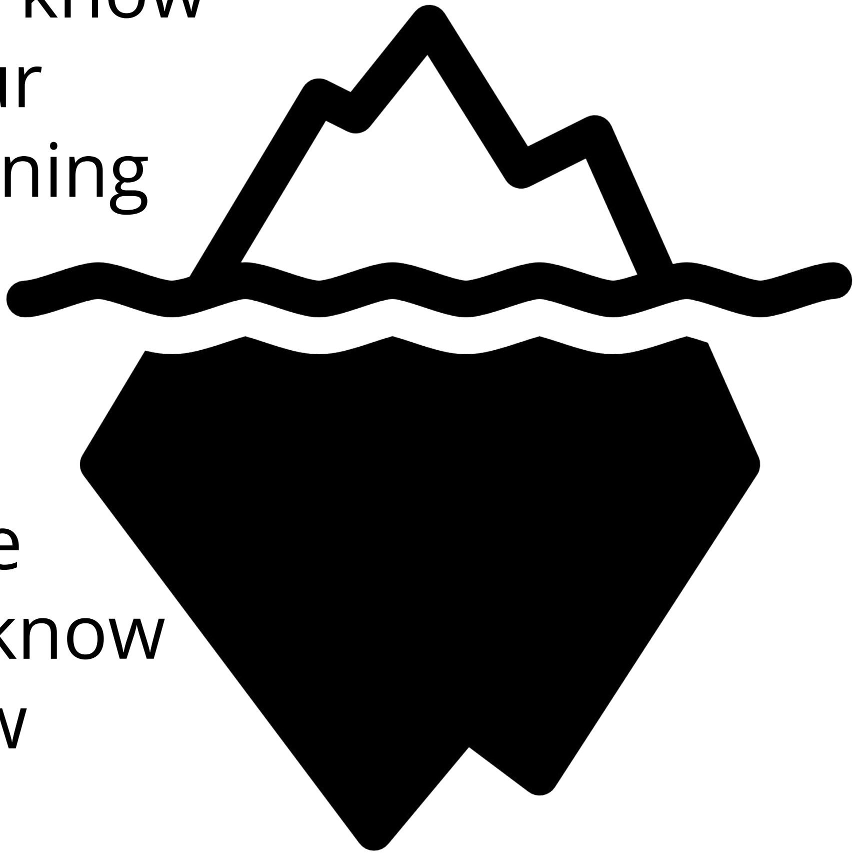
- Model Tracing: Follows each student's thinking steps
  => context-sensitive instruction
- Knowledge Tracing: Assesses student's knowledge growth
  => individualized activity selection and pacing



Koedinger et al. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*.

Intuition-based design *lacks information*

What we know
about our
own learning



What we
do **not** know
we know

"experts are not fully
aware of about **70%** of
their own decisions"
- Richard Clark in
Cognitive Task Analysis

Data breaks illusions

Data => Insight => Adaptation that works for all students

Which is harder for algebra students?

Story Problem

As a waiter, Ted gets $6 per hour. One night he made $66 in tips and earned a total of $81.90. How many hours did Ted work?

Word Problem

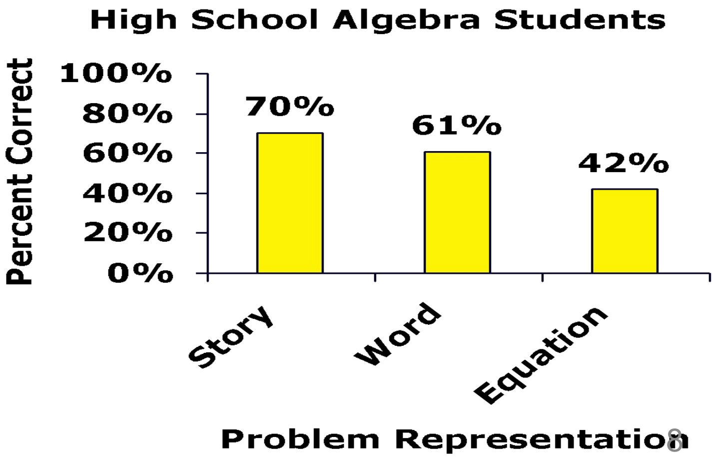
Starting with some number, if I multiply it by 6 and then add 66, I get 81.90. What number did I start with?

Equation

x * 6 + 66 = 81.90

Math educators say:
story or word is hardest

Students:
equations
are hardest



Expert blind spot!
Algebra teachers, especially, incorrectly think equations are easy

Koedinger & Nathan (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Learning Science.*

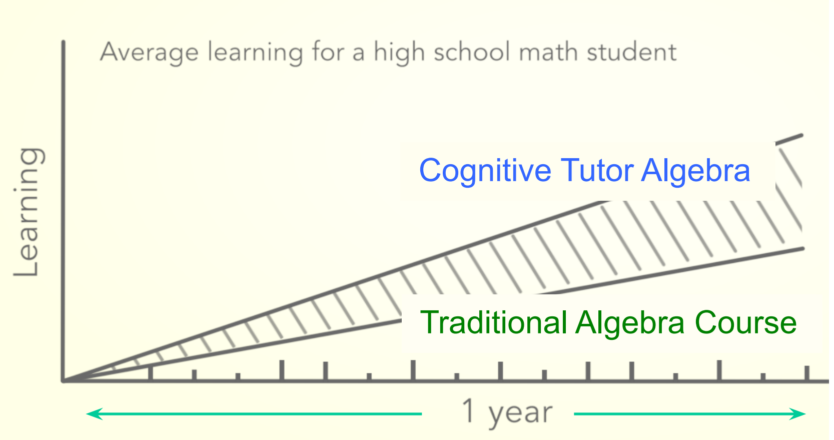
Interactive Practice Works at Scale

K12 Math Cognitive Tutors

- Widely used
  - ~500K students per year
  - ~80 minutes per week
- *2x better learning*

College Online Courses

- Widely used
  - ~50K-1M students per year
  - ~30 courses at 1K colleges
- *2x faster learning*



Pane et al. (2013). Effectiveness of Cognitive Tutor Algebra I at Scale. Santa Monica, CA: RAND Corp.

Carnegie Mellon University

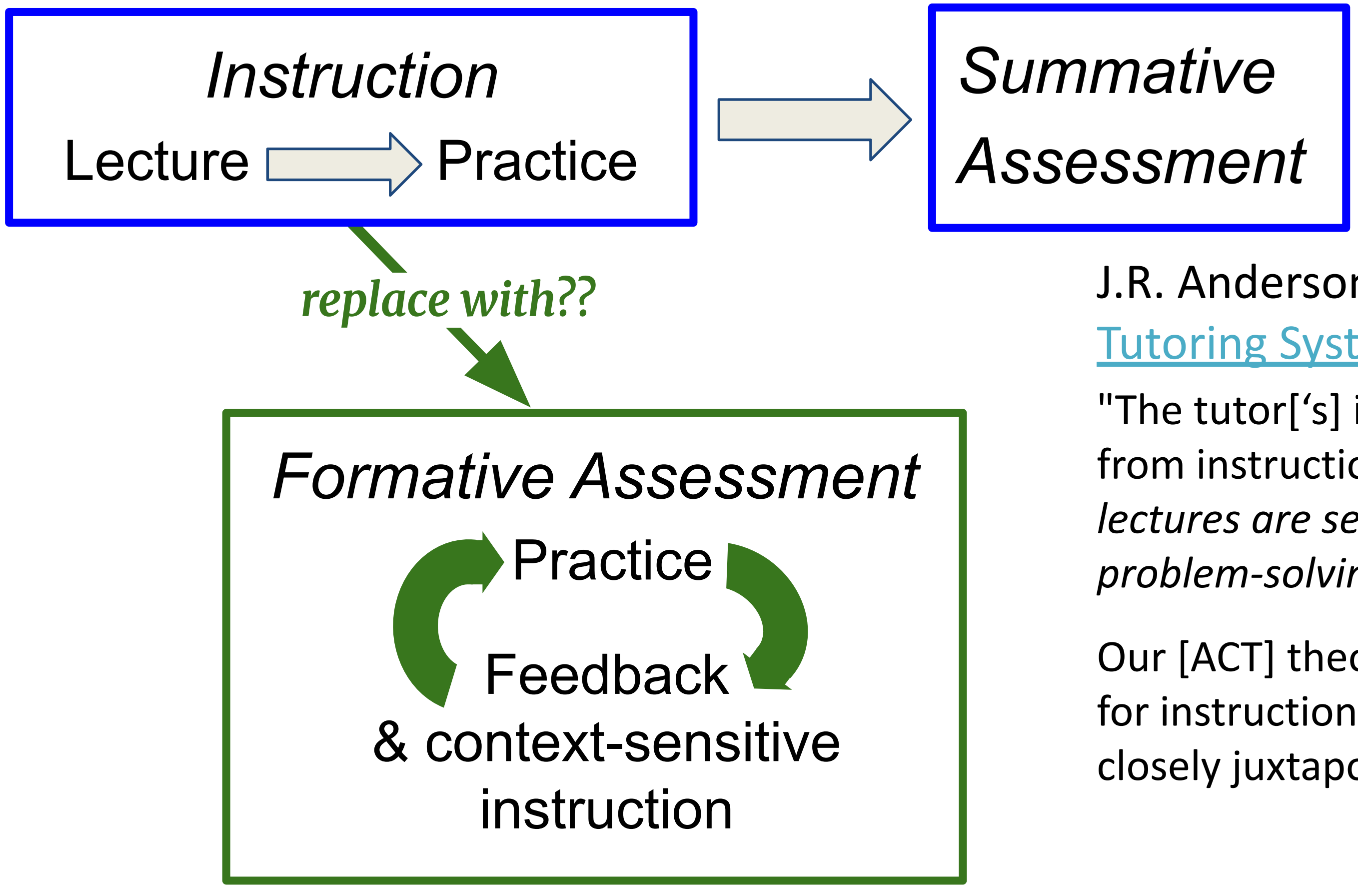
Open Learning Initiative

Transforming higher education through the science of learning.

Lovett et al. (2008). Measuring the effectiveness of the OLI learning course in accelerating student learning. *Interactive Media in Education.*

ITS as formative assessment & a rethinking of educational sequencing

Most current education:



J.R. Anderson (1985). [Intelligent Tutoring Systems](). *Science*.

"The tutor['s] instruction mode differs from instructional modes in which *lectures are separate from problem-solving* ….

Our [ACT] theory implies that it is critical for instruction and problem-solving to be closely juxtaposed."

Learning by doing

> "Give the pupils something to do, not something to learn; and the doing is of such a nature as to demand thinking; learning naturally results" – John Dewey

Lots of evidence with different catch phrases:

- Testing effect

  Roediger & Karpicke (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psych Science*

- Deliberate practice

  Ericsson et al (1993). The role of deliberate practice in the acquisition of expert performance. *Psych Rev*
  Duckworth et al (2011). Deliberate practice spells success: Why grittier competitors triumph at the National Spelling Bee. *Soc Psych & Pers Sci*

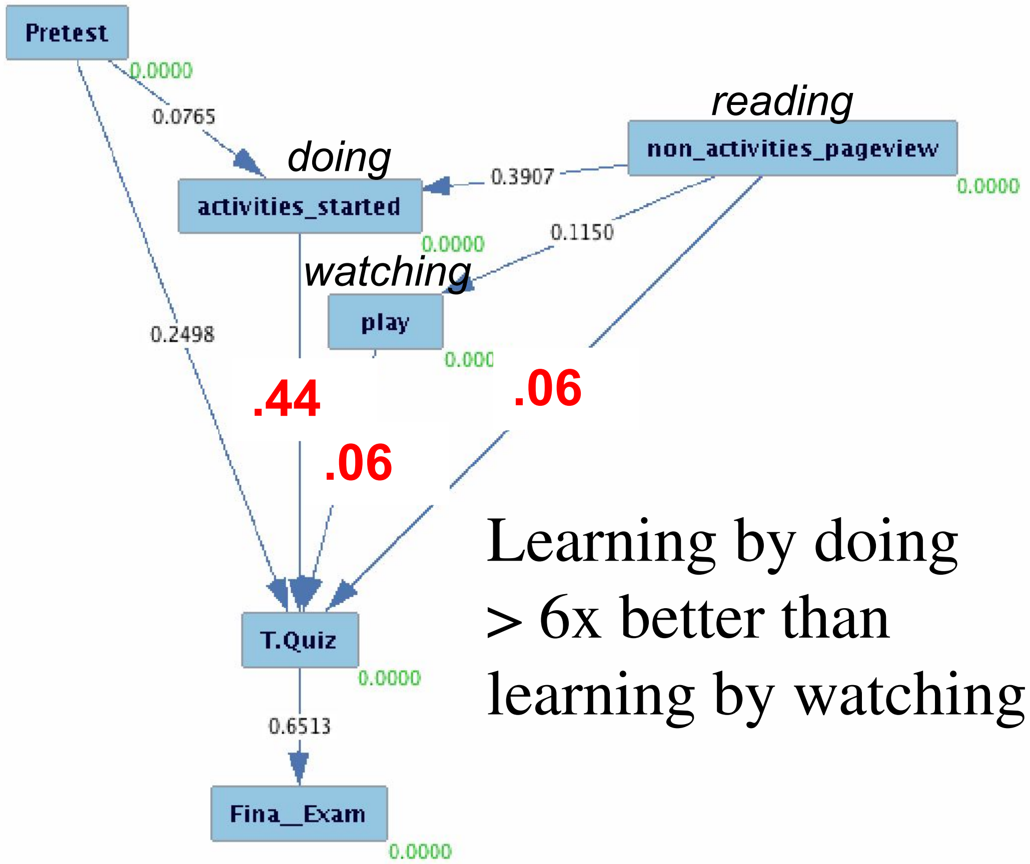
- Active learning

  Freeman et al. (2014). Active learning increases student performance in science, engineering and mathematics. *PNAS*
  Deslauriers et al (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *PNAS*

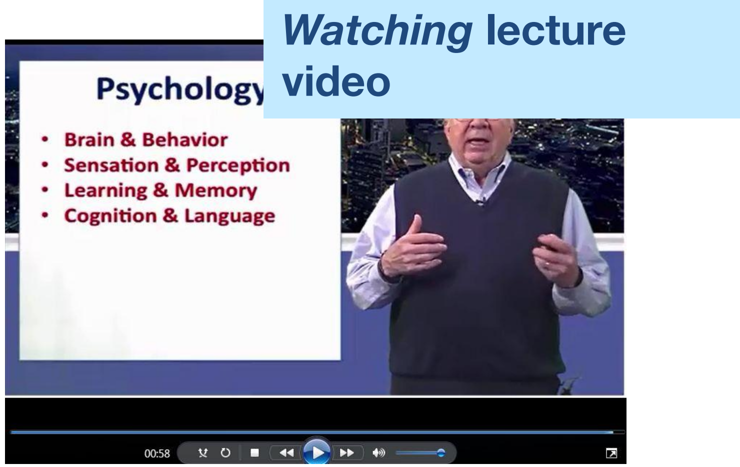
- Doer effect

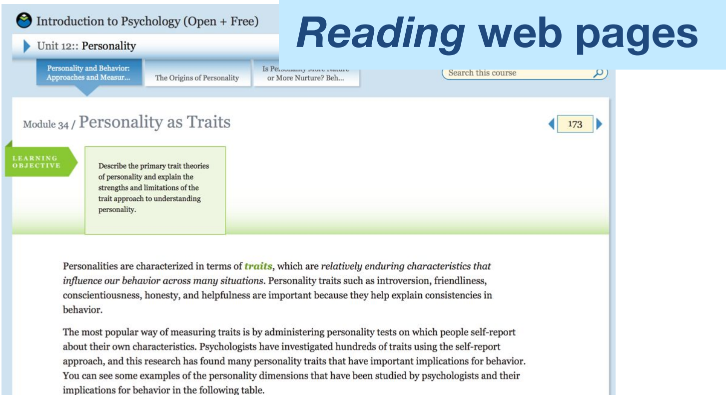
  …

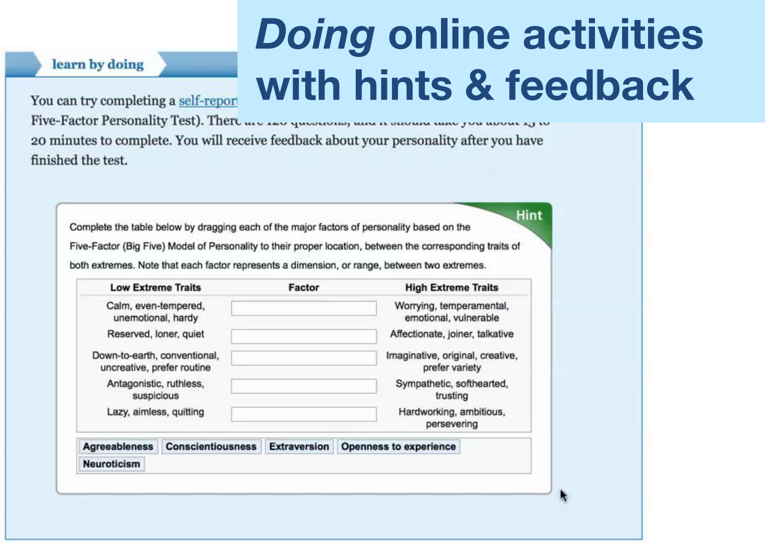
Doer Effect



Learning by doing > 6x better than learning by watching!







Koedinger et al. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. *Learning at Scale.*

Doer effect found in other courses & by other researchers

	Doing to reading effect ratio	
	Quizzes	**Final Grade**
Info Systems	5.2	2.2
Biology	5.0	3.1
Statistics	∞	16.4
Psych	8.5	7.7
Psy MOOC	6.8	4.8

Van Campenhout, Jerome, Dittel & Johnson (2023). The Doer Effect at Scale: Investigating Correlation and Causation Across Seven Courses. In *Proceedings of LAK*. https://doi.org/10.1145/3576050.3576103

N = 12K students
Doer effect range: 2.2x to 16x   Median: 6x

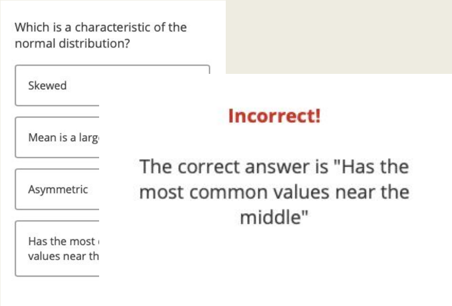
Koedinger, McLaughlin , Jia, & Bier (2016).
Is the Doer Effect a Causal Relationship?
In *Proceedings of LAK*.

Doer Effect Experiment

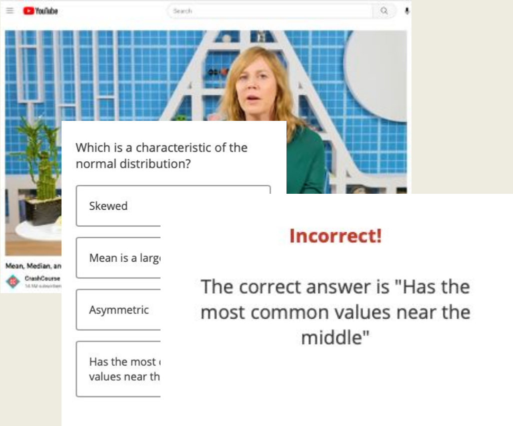
Learning phase: Central tendency

OR

OR



OR



No instruction

Instructional video

Practice with feedback

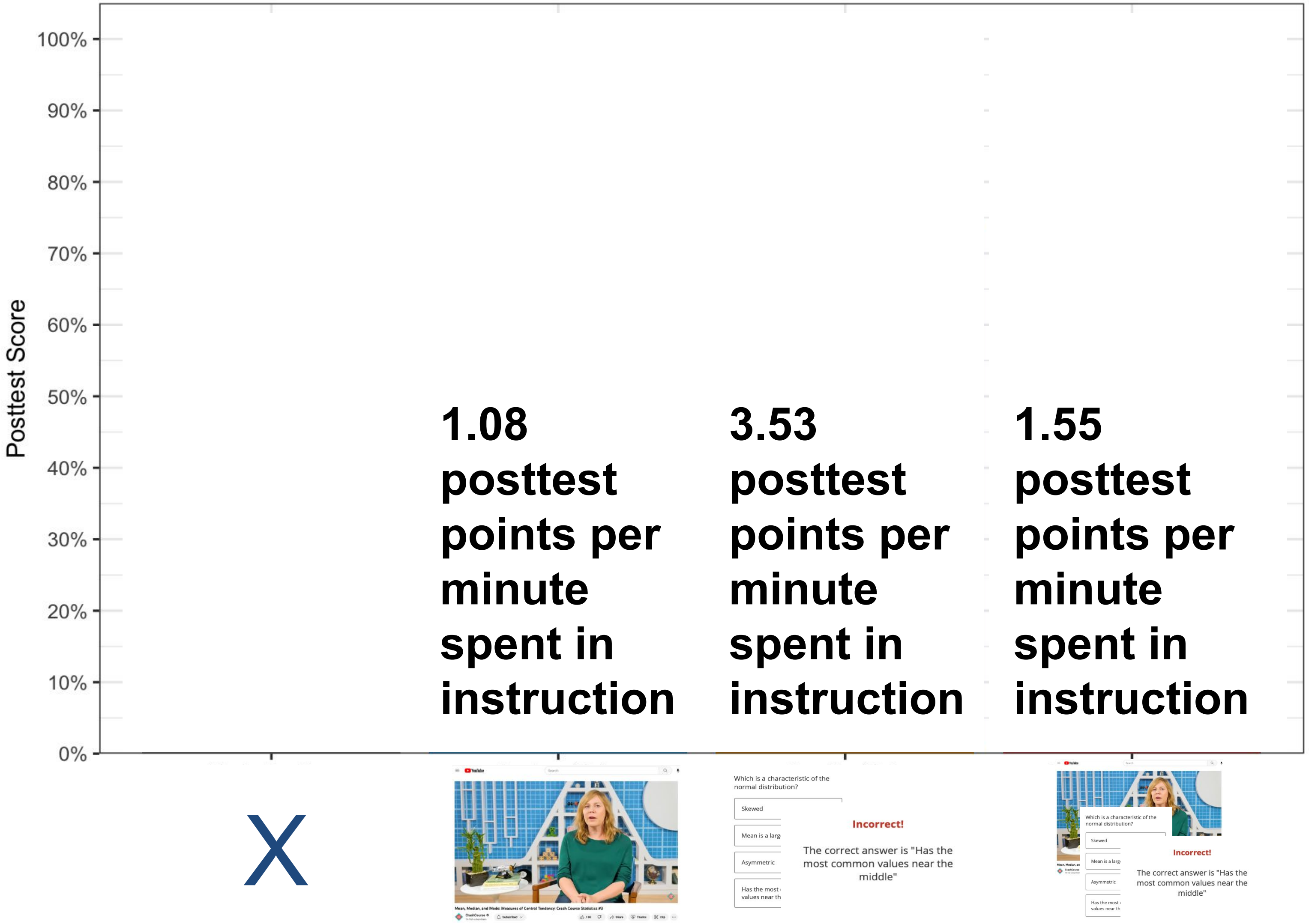
Instructional video
& practice with feedback

Post assessment

Near & far transfer questions about central tendency

Asher, Sana, Koedinger, & Carvalho (2025). Practice with feedback versus lecture: Consequences for learning, efficiency, and motivation. J of Applied Research in Memory and Cognition.

Practice with feedback produces more efficient learning

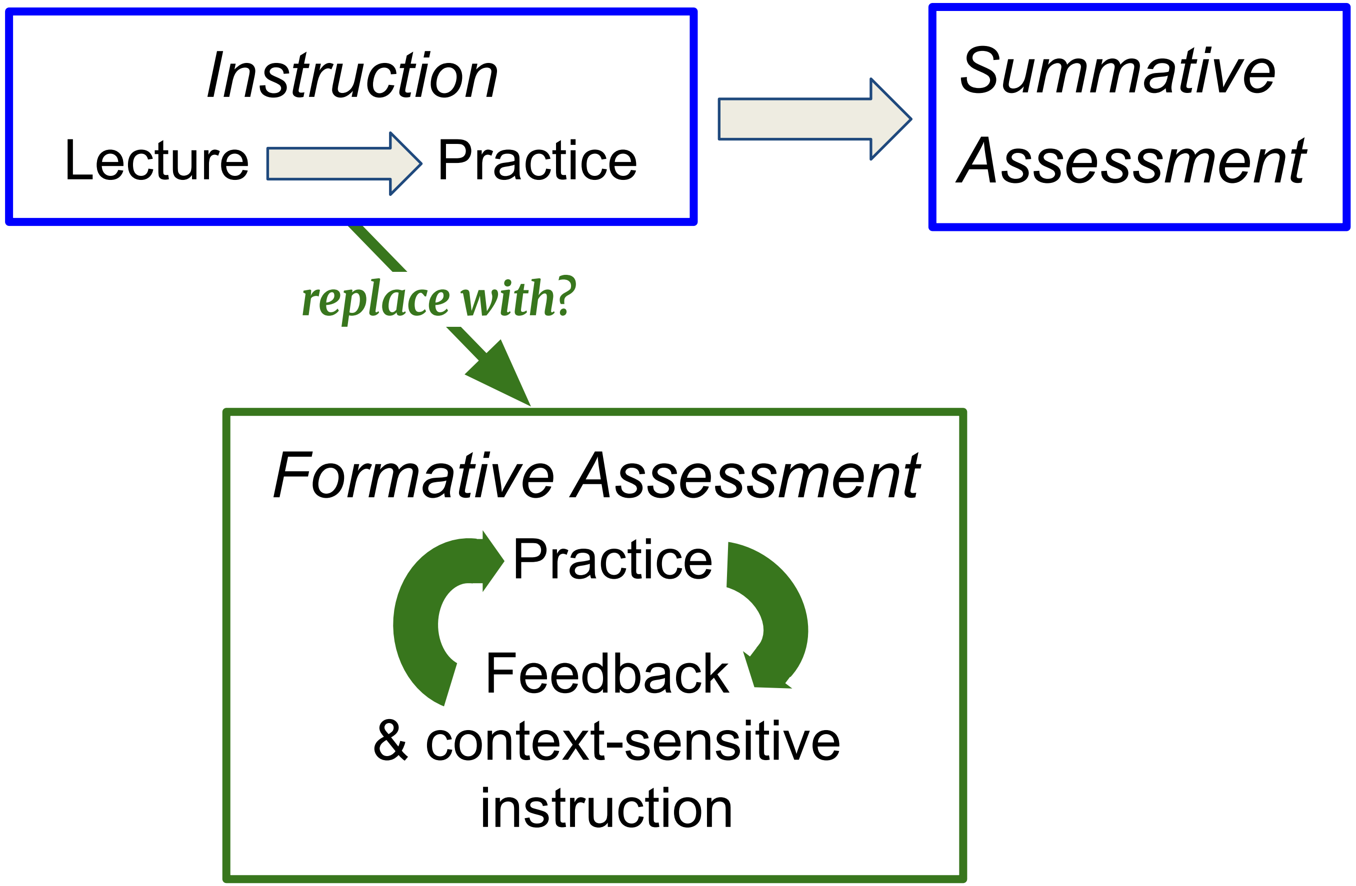


Practice with feedback
3x better than lecture
2.4x better than combined

Asher et al (2025). Practice with feedback versus lecture. J of Applied Research in Memory & Cognition.

More direct experimental evidence re rethinking educational sequencing

Most current education:



Overview

Interactive learning by doing & online tutors

What student differences account for learning outcome differences?

Koedinger, Carvalho, Liu, McLaughlin (2023). An Astonishing Regularity in Student Learning Rate. *PNAS*.

Formative Assessment instead of Summative Assessment

Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*.

Can hybrid human-AI tutoring enhance educational equity?

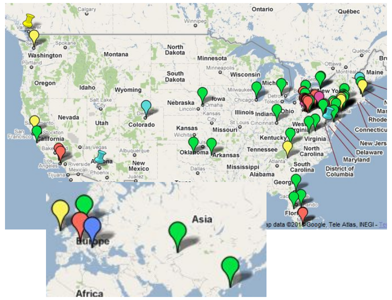
Chine et al (2022). Educational Equity Through Combined Human-AI Personalization. AIEd Conference.

Holstein et al (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

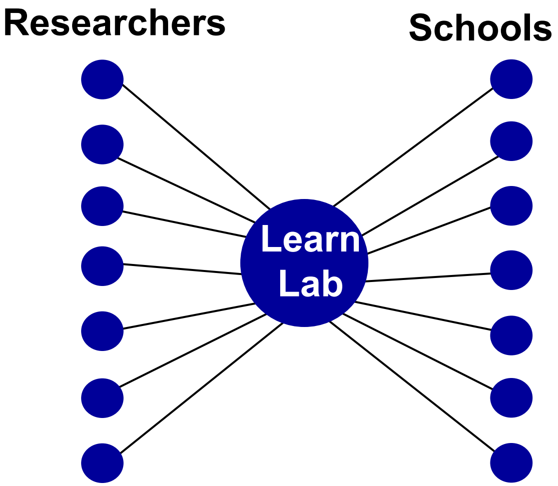
Tech infrastructure advances science

Ed tech + wide use = "Basic research *at scale*"

+



= LearnLab



$47M, 2004-15

> 360 *in vivo* experiments

> 4000 ed tech data sets in DataShop

PSLC DATASHOP
a data analysis service for the learning science community

Lessons from 360 *in vivo* experiments
=> KLI Framework

Ed

Ed

Psych

Psych

Instructional Principles (simpler on bottom)	Chinese vocab	French articles	English articles	Algebra eq	Geometry rules	Chemistry rules	Help seeking skls	Physics prinpls	Chem models	Pressure concept
Accountable Talk									+ +	+
Collaboration				0						+
Self-explanation			-	+	+			+		
Worked examples	-		0	+	+ + +	+				
Diagram coordination					+				+	
Feature Focusing	+	+ + +								
Feedback		+		+			+			
Optimal Scheduling	+				+					
	Facts			Rules				Principles		

Knowledge Components
(simpler on left)

1. Ideal instruction ***depends*** on knowledge goals

2. Because ***different learning processes*** are at work

3. Makes sense of competing recommendations

Koedinger, Corbett, & Perfetti (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*.

Do students learn at different rates?

Yes:

- "high-ability learners learn at a more rapid rate than other students"
  - National Academy of Sciences report

Not so much:

- "most students become very similar with regard to ... rate of learning ... when provided with favorable learning conditions"
  - Benjamin Bloom

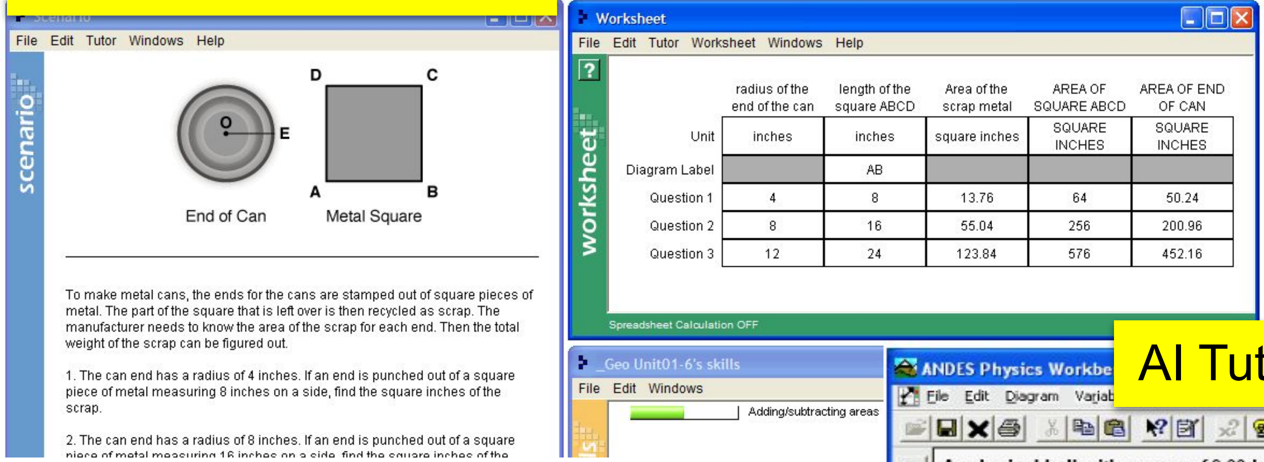
Notes

- No doubt: students achieve differently
- But: do they start differently or have more opportunities
- We control for these to measure learning rate

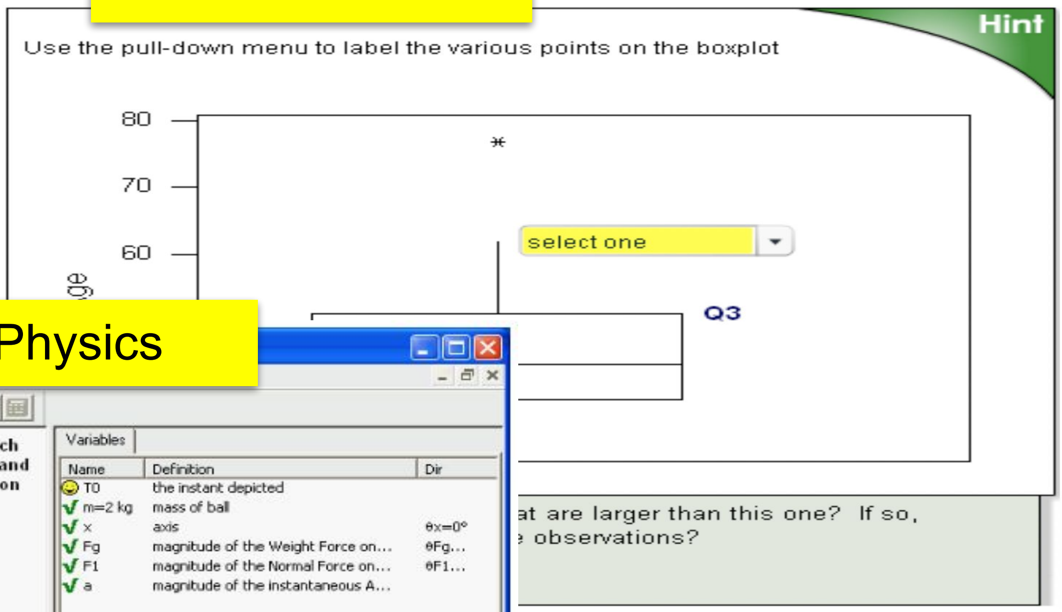
Koedinger, Carvalho, Liu, McLauglin (2023).
An Astonishing Regularity in Student Learning Rate.

Interactive learning in 27 datasets from K12 & College courses using AI Tutors, Online Courses, Ed Games

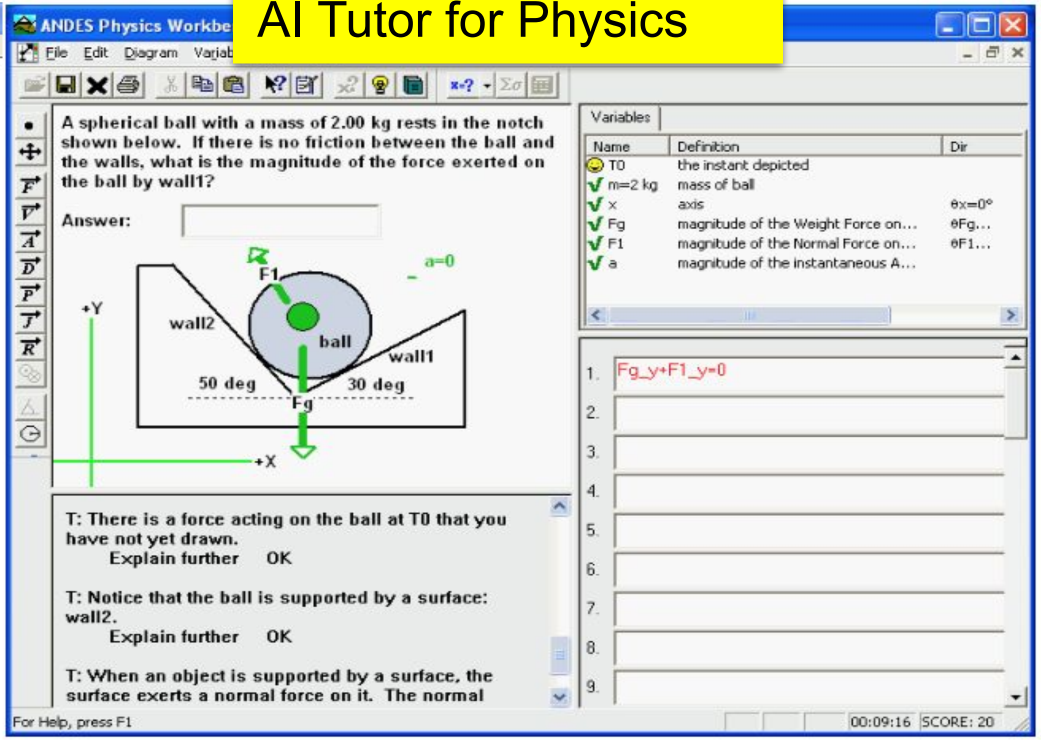
AI Tutor for Geometry



Statistics Online



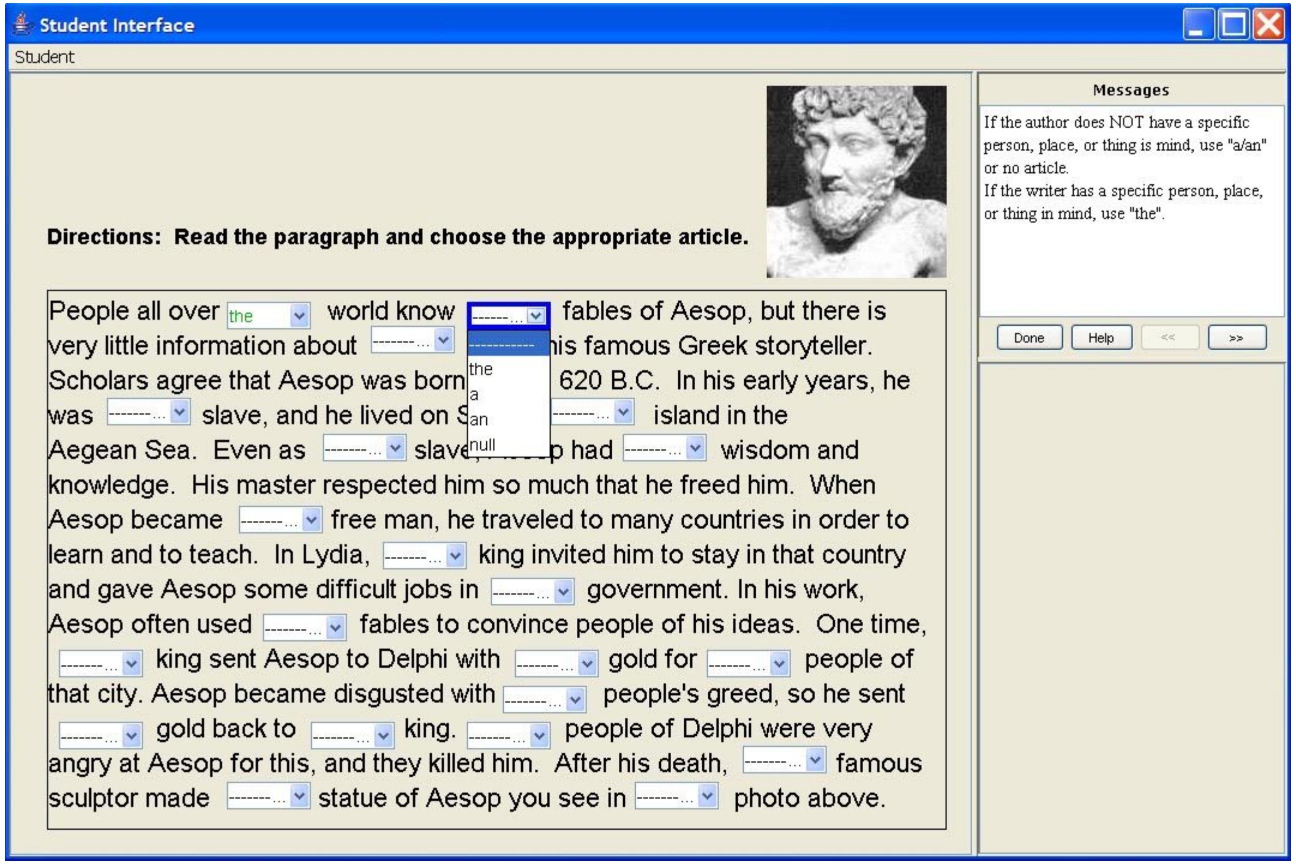
AI Tutor for Physics



Battleship Numberline Game

Interactive learning in 27 datasets from K12 & College courses using AI Tutors, Online Courses, Ed Games

English Articles



Chinese Vocabulary

Chinese	*English*?
Lao3shi1	$\boxed{\text{Teacher}}$

Learning Curves Require a Cognitive Model

- If we compare observations on **matched tasks**, we *can observe learning improvement*
- If we compare observations on **unmatched tasks**, we *cannot*
- LearnLab produced empirical methods to group tasks involving the same knowledge components
  - Such a grouping of empirically matched tasks is a cognitive model

Q matrix as a way to match tasks & explain task (or item) difficulty

- LTTM is an “item explanatory” generalization of Rasch model (IRT(
- Wilson & DeBoeck, 2004

	$\eta_{pi} =$			
Model	Person part	Item part	Random effect	Model type
Rasch model	θ_p	$-\beta_i$	$\theta_p \sim N(0, \sigma_\theta^2)$	Doubly descriptive
Latent reg Rasch model	$\sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p$	$-\beta_i$	$\varepsilon_p \sim N(0, \sigma_\varepsilon^2)$	Person explanatory
LLTM	θ_p	$-\sum_{k=0}^{K} \beta_k X_{ik}$	$\theta_p \sim N(0, \sigma_\theta^2)$	Item explanatory
Latent reg LLTM	$\sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p$	$-\sum_{k=0}^{K} \beta_k X_{ik}$	$\varepsilon_p \sim N(0, \sigma_\varepsilon^2)$	Doubly explanatory

Accurate cognitive model of transfer of learning needed

Cognitive models specify **Knowledge Components (KCs)** needed to succeed in **tasks**

- q_{jk} matrix maps tasks j to KCs k

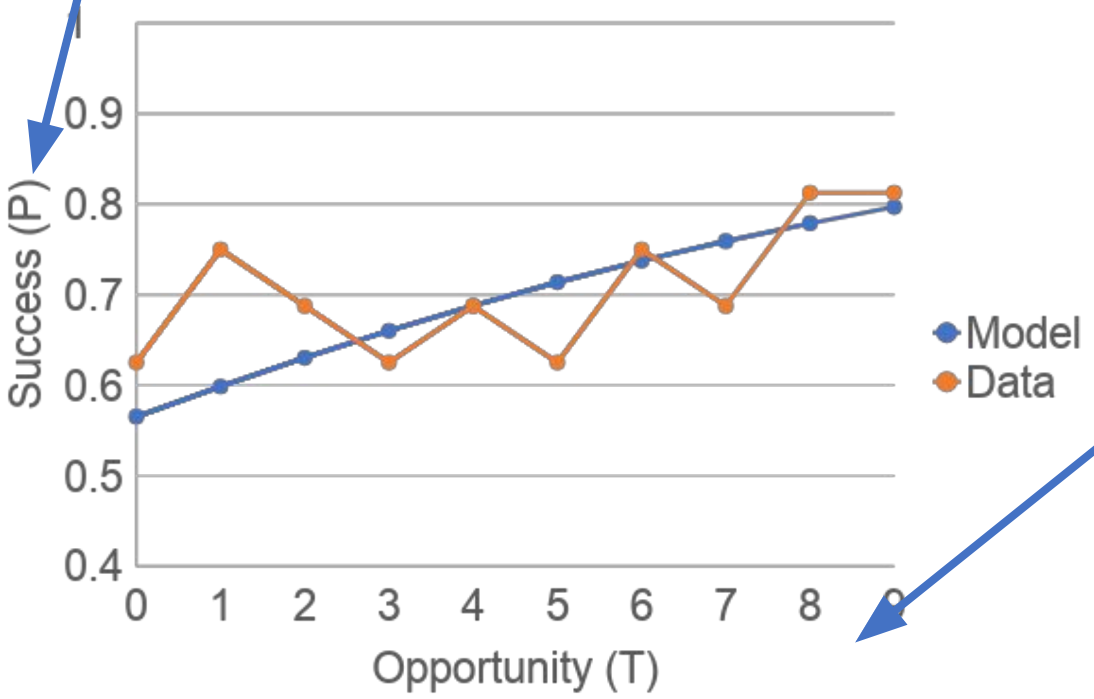
Tasks j (Observed Problem Steps)	Q0	Q1	
	Arith	Mult	Sub
2*8-30 => 16-30	1	1	0
16-30 => -14	1	0	1
30-2*8 => 30-16	1	1	0
30-16 => 14	1	0	1
10-3*7 => 10-21	1	1	0
10-21 => -11	1	0	1

Which Q is right?
Best fit to curve

A poor Q matrix produces poor fit to learning curves

Learning Curves: Success grows with opportunity

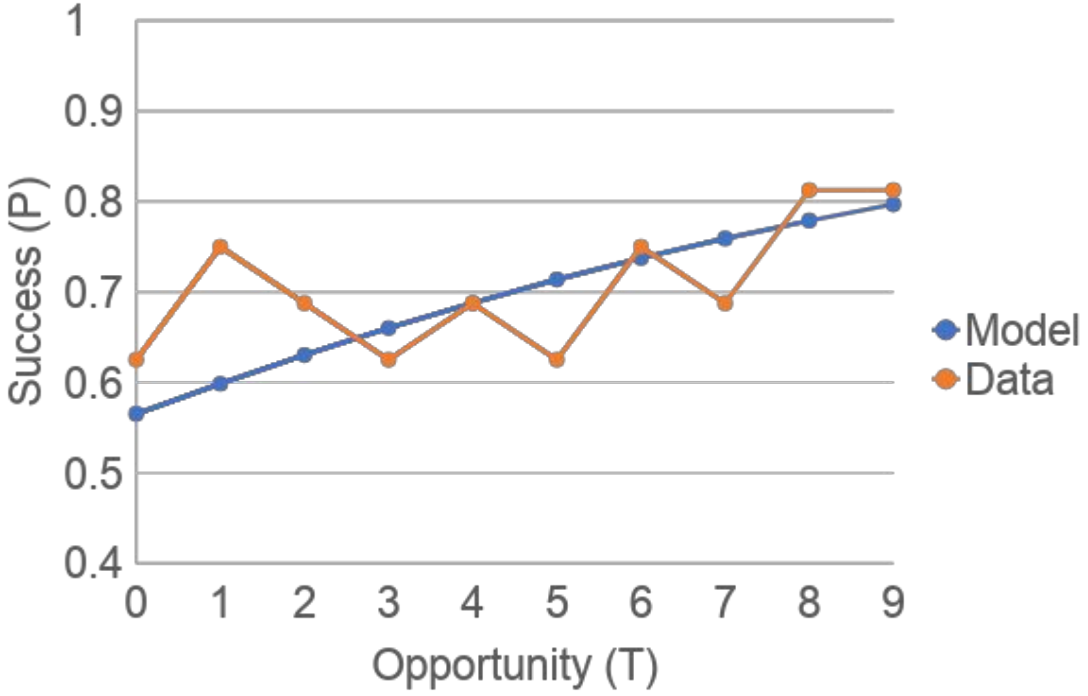
Success (p) ~ Initial-Knowledge + Learning-Rate * Opportunities

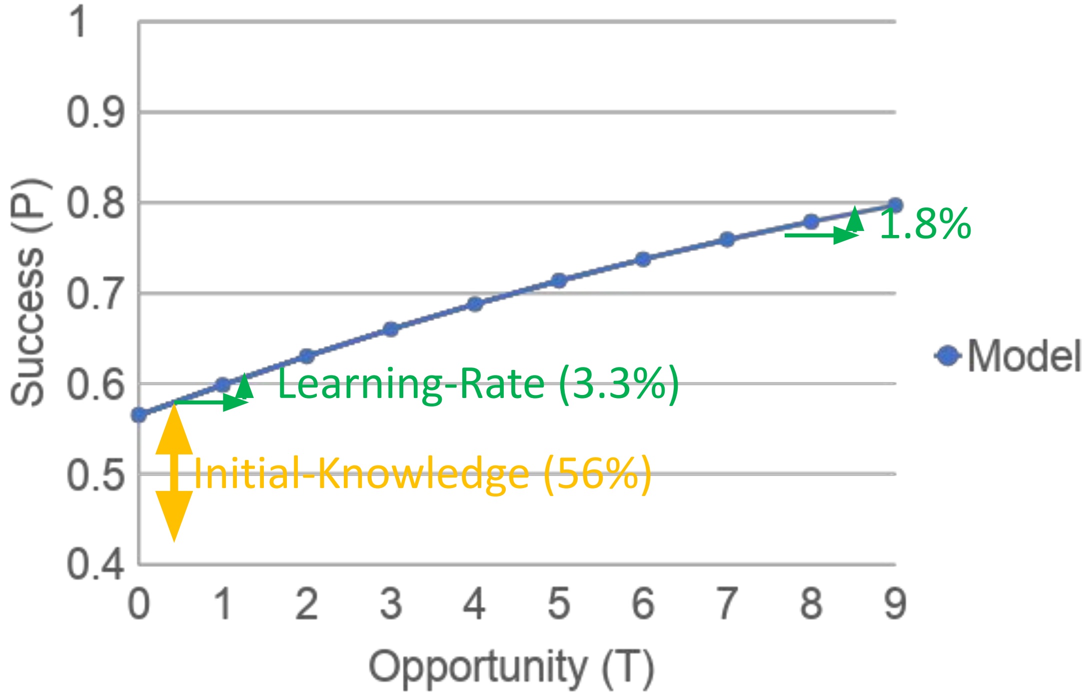


Learning Curve modeling basics

Success (p) ~ Initial-Knowledge + Learning-Rate * Opportunities

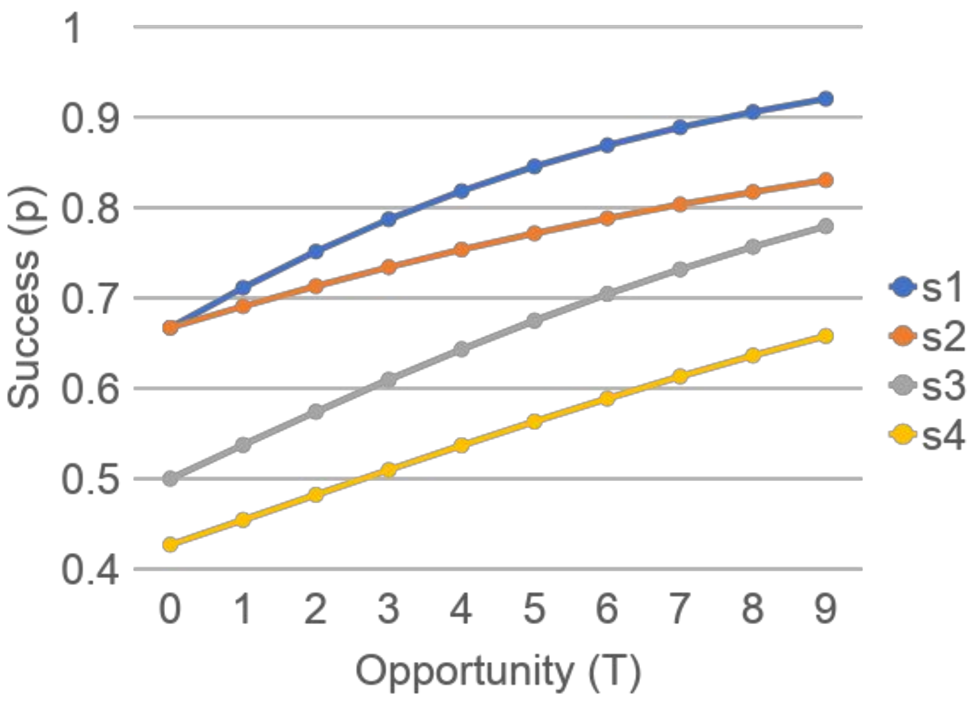
y ~ b + mx

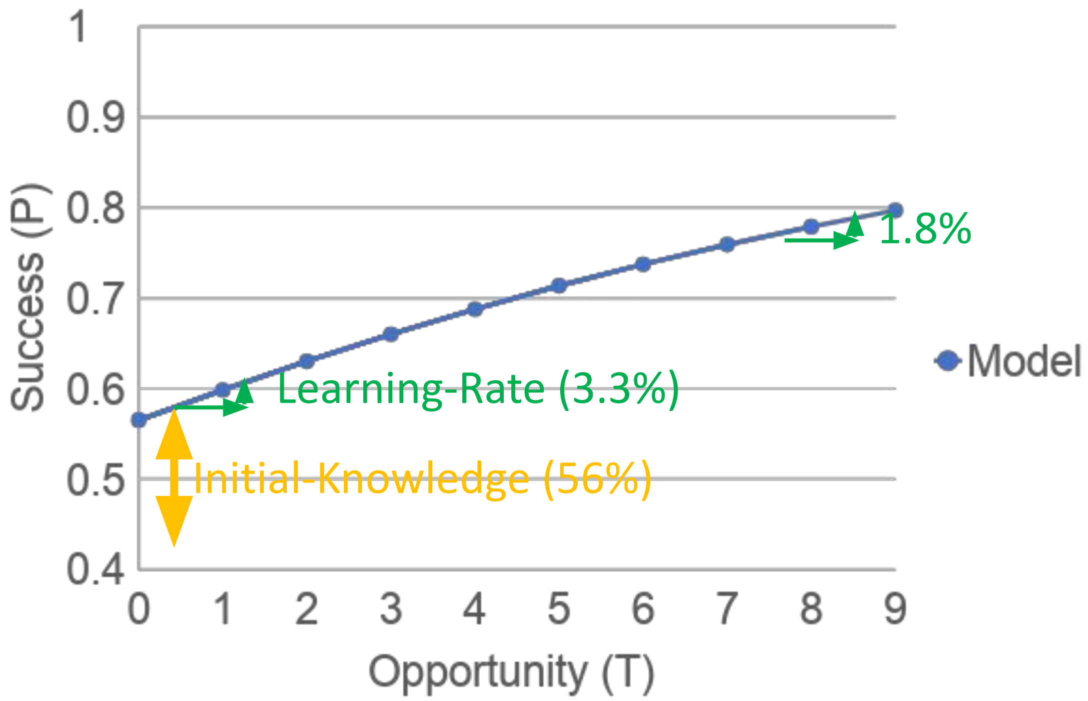




Learning Curves by student (i)

Success (p_i) ~ Initial-Knowledge$_i$ + Learning-Rate$_i$ * Opportunities$_i$



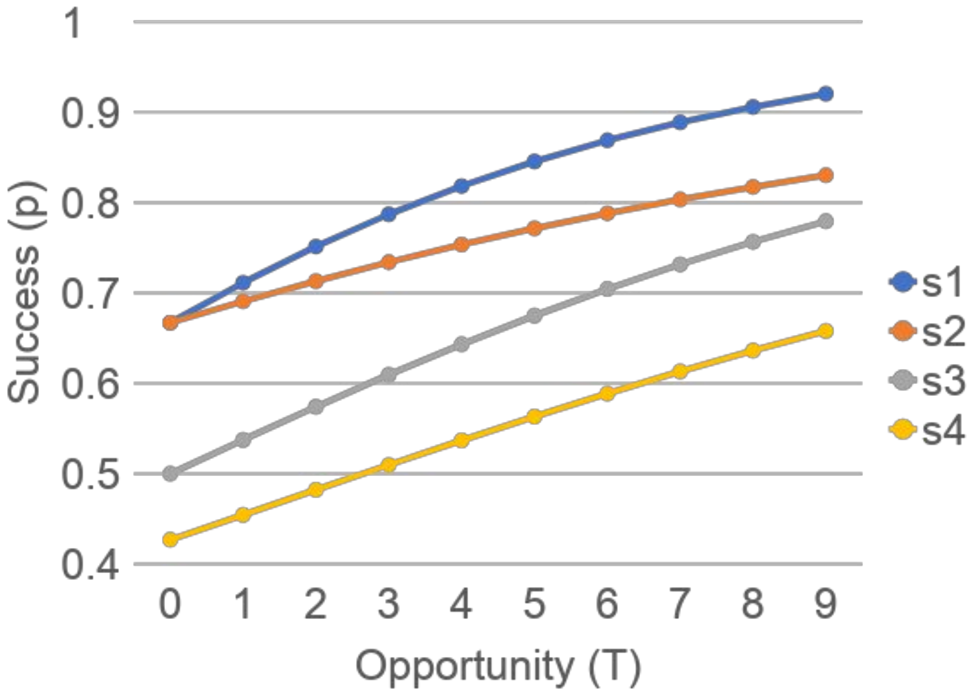


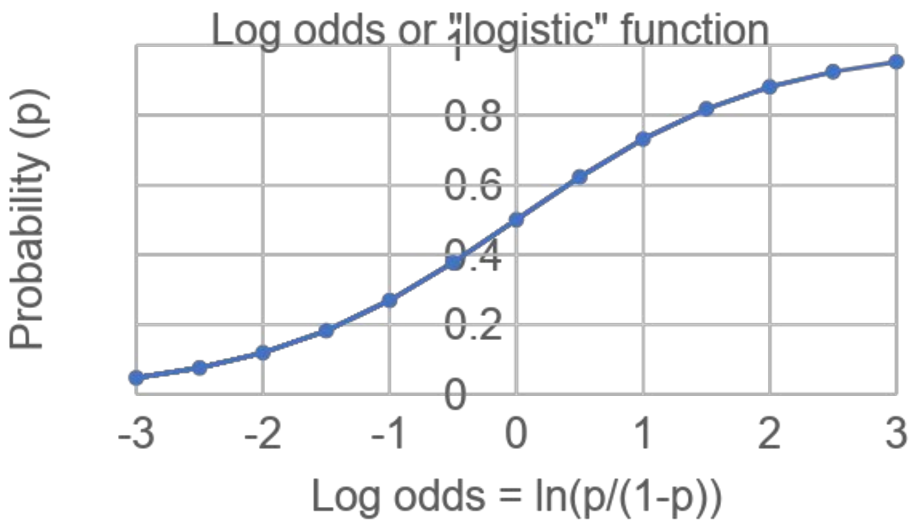
Learning Curves: Transforming to log odds

Success (p_i) ~ Initial-Knowledge$_i$ + Learning-Rate$_i$ * Opportunities$_i$

$\ln(p_i/(1-p_i))$ = Initial-Knowledge$_i$ + Learning-Rate$_i$ * Opportunities$_i$

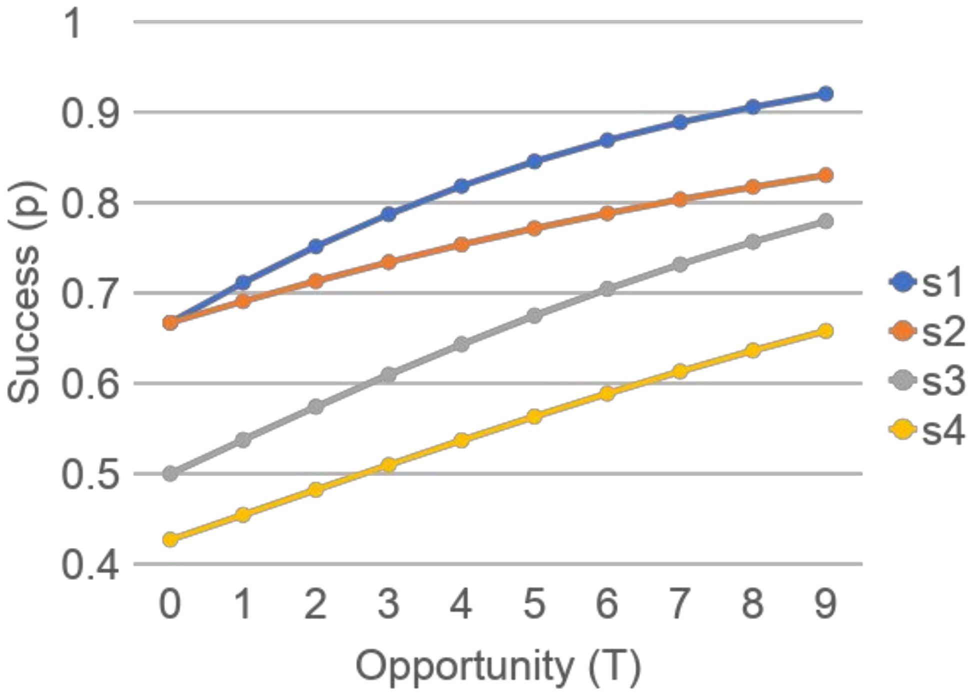
Logistic regression

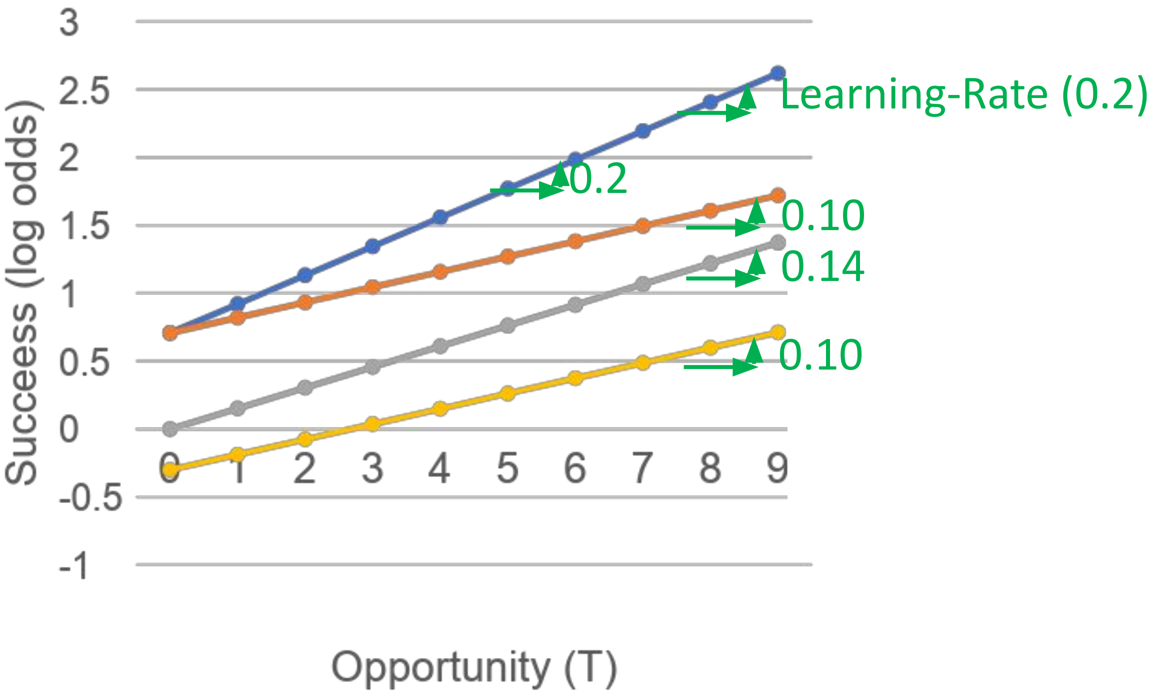




Learning Curves: Linear increase per opportunity

$\ln(p_i/(1-p_i)) = \text{Initial-Knowledge}_i + \text{Learning-Rate}_i * \text{Opportunities}_i$





All 27 datasets used a KC model (Q matrix) that was improved through *iterative learning engineering*

Tasks j (Observed Problem Steps)	Q0	Q1		Q2				Q3 = Item model					
	Arith	Mult	Sub	Mult LR	Mult OR	Sub+	Sub-	I1	I2	I3	I4	I5	I6
2*8-30 => 16-30	1	1	0	1	0	0	0	1	0	0	0	0	0
16-30 => -14	1	0	1	0	0	0	1	0	1	0	0	0	0
30-2*8 => 30-16	1	1	0	0	1	0	0	0	0	1	0	0	0
30-16 => 14	1	0	1	0	0	1	0	0	0	0	1	0	0
10-3*7 => 10-21	1	1	0	0	1	0	0	0	0	0	0	1	0
10-21 => -11	1	0	1	0	0	0	1	0	0	0	0	0	1

Chosen Q matrix for each dataset had to fit better than the extremes Q0 and Q3

Integrating Cognitive Model into KC model

$\ln(p_i/(1-p_i))$ = Initial-Knowledge$_i$ + Learning-Rate$_i$ * Opportunities$_i$

$\ln(p_i/(1-p_i))$ = S_initial + KC_initial + …

$$ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \theta_i + \sum_{k=1}^{K} q_{jk}\beta_k + \dots$$

Cognitive model explains why tasks (j) are hard in terms of KCs (k)

Generalization of item response theory (Wilson & DeBoeck, 2004)
Appears *without student rate* in Spada &McGaw (1985), Draney, Pirolli, & Wilson (1995), Cen et al (2006)
Liu et al found highly correlated student initial & rate estimates using a Bayesian Hidden Markov model (BKT)

Integrating Cognitive Model into KC model

$\ln(p_i/(1-p_i))$ = Initial-Knowledge$_i$ + Learning-Rate$_i$ * Opportunities$_i$

$\ln(p_i/(1-p_i))$ = S_initial + KC_initial + (S_rate + KC_rate)* Opportunities$_{ik}$

$$ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \theta_i + \sum_{k=1}^{K} q_{jk}\beta_k + (\delta_i + \sum_{k=1}^{K} q_{jk}\gamma_k)T_{ik}$$

Cognitive model explains why tasks (j) are hard in terms of KCs (k)

Generalization of item response theory (Wilson & DeBoeck, 2004)
Appears *without student rate* in Spada &McGaw (1985), Draney et al (1995), Cen et al (2006)
Liu et al found highly correlated student initial & rate estimates using a Bayesian Hidden Markov model (BKT)

Four key results across 27 data sets

- Student learning rate factor improves prediction but not always
  - AIC: 21 of 27   BIC: 15 of 27
  - Student learning rate variation is detectable
- Typical students starts at about 65% correctness
  - Needs about 7 practice repetitions to reach mastery at 80%

Learning Theory: Practice is crucial

Limited accuracy after up-front verbal instruction =>

Human learning is

- not simply about
  explicit processing, encoding, & retrieval of verbal instruction
- as much or more about
  implicit or nonverbal learning-by-doing
  in varied task settings with interactive feedback available

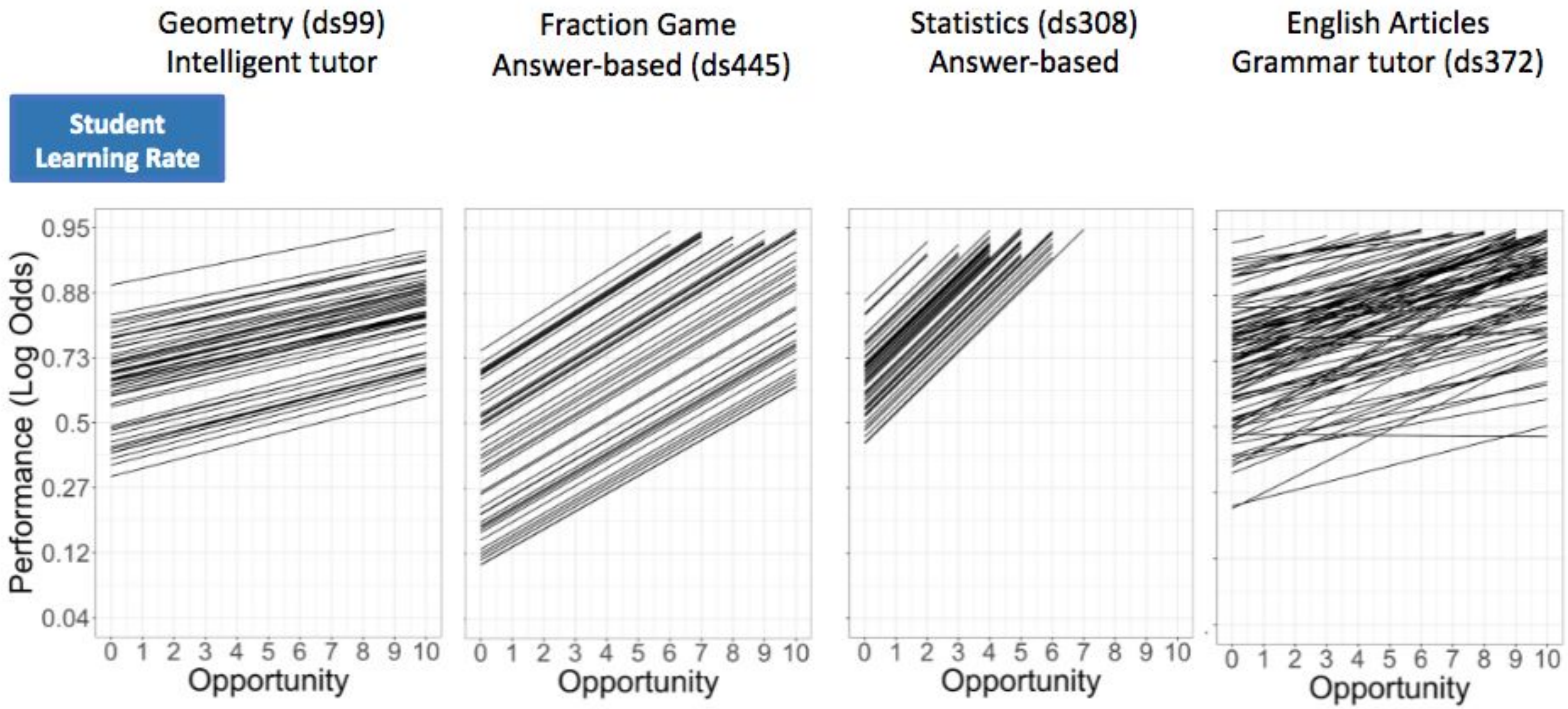
Four key results across 27 data sets

- Student learning rate factor improves prediction but not always
  - AIC: 21 of 27   BIC: 15 of 27
  - Student learning rate variation is detectable
- Typical students starts at about 65% correctness
  - Needs about 7 practice repetitions to reach mastery at 80%
- Students **vary substantially** in *initial knowledge*
- Students are **astonishingly similar** in *learning rate*

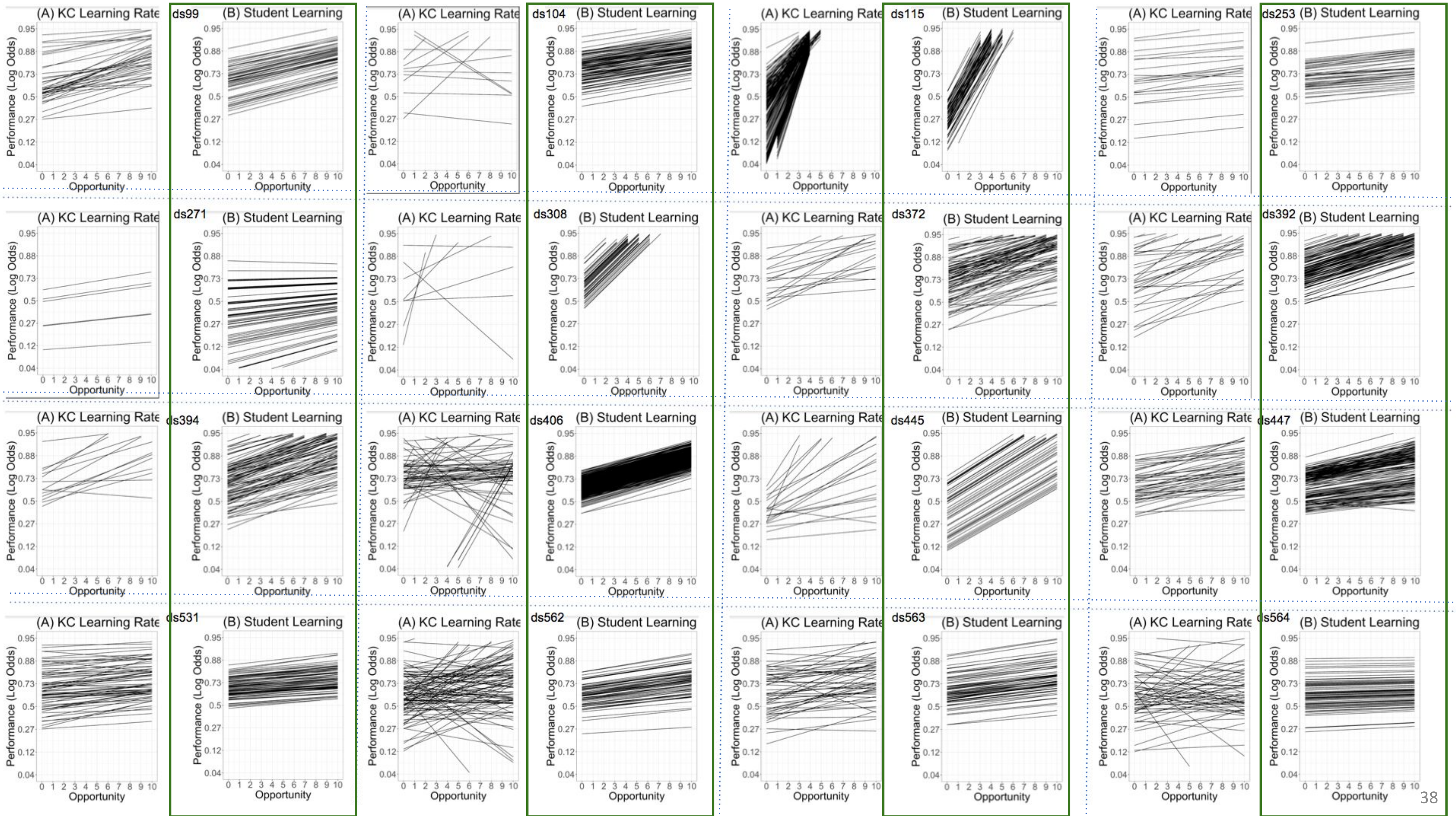
Student initial performance & learning rate variation in sample datasets

Parallel lines:
small variation in learning rate

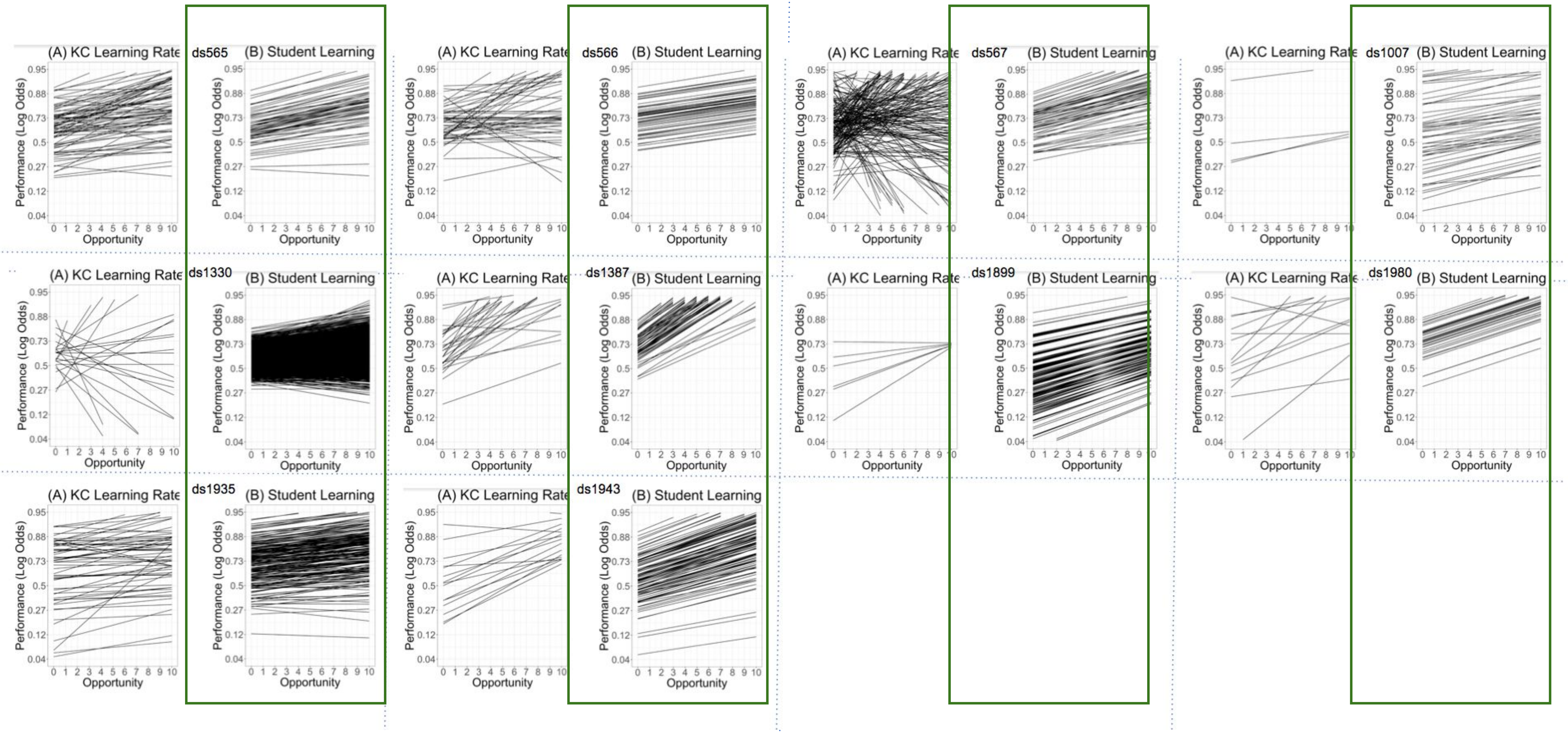
Intercepts:
large variation in initial performance



Koedinger, Carvalho, Liu, McLaughlin (2023). An Astonishing Regularity in Student Learning Rate. *PNAS*.



ds99
ds104
ds115
ds253
ds271
ds308
ds372
ds392
ds394
ds406
ds445
ds447
ds531
ds562
ds563
ds564
(A) KC Learning Rate
(B) Student Learning
Performance (Log Odds)
0.95
0.88
0.73
0.5
0.27
0.12
0.04
0 1 2 3 4 5 6 7 8 9 10
Opportunity


ds565
(A) KC Learning Rate
(B) Student Learning
ds566
ds567
ds1007
ds1330
ds1387
ds1899
ds1980
ds1935
ds1943
Performance (Log Odds)
Opportunity

Initial variation requires 10x more opportunities to catch up than rate variation

	Initial knowledge		Learning rate	
Percentile	Initial % correct	Opp to reach 80% mastery	Improvement % correct*	Opp to reach 80% mastery
25	55.21 (15.84)	13.13 (19.52)	1.70 (3.80)	7.89 (22.41)
50	66.05 (12.91)	6.54 (14.10)	2.25 (4.02)	7.27 (14.18)
75	75.17 (10.45)	3.66 (8.22)	2.56 (4.12)	6.94 (10.91)

<u>Initial knowledge</u>

55% to 75% crt

10 opportunities needed to catch up

<u>Learning rate</u>

1.7% to 2.6%/opp

1 opportunity needed to catch up

What accounts for learning outcome differences?

*Incoming differences by an **order of magnitude** over learning rate diffs*

Promising Conclusion

Given

favorable learning conditions for deliberate practice

learner invests effort in learning opportunities

anyone can learn anything they want!

Original goal:
Identify high ability learners & understand their characteristics

Surprising outcome:
Highly similar rates of learning across students

Message for young researchers & innovators:
When things do *not* work as expected,
you may be onto something interesting. Stick with it!

Koedinger, Carvalho, Liu, McLaughlin (2023). An Astonishing Regularity in Student Learning Rate. *PNAS*.

Overview

Interactive learning by doing & online tutors

What student differences account for learning outcome differences?

Koedinger, Carvalho, Liu, McLaughlin (2023). An Astonishing Regularity in Student Learning Rate. *PNAS*.

Formative Assessment instead of Summative Assessment

Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*.

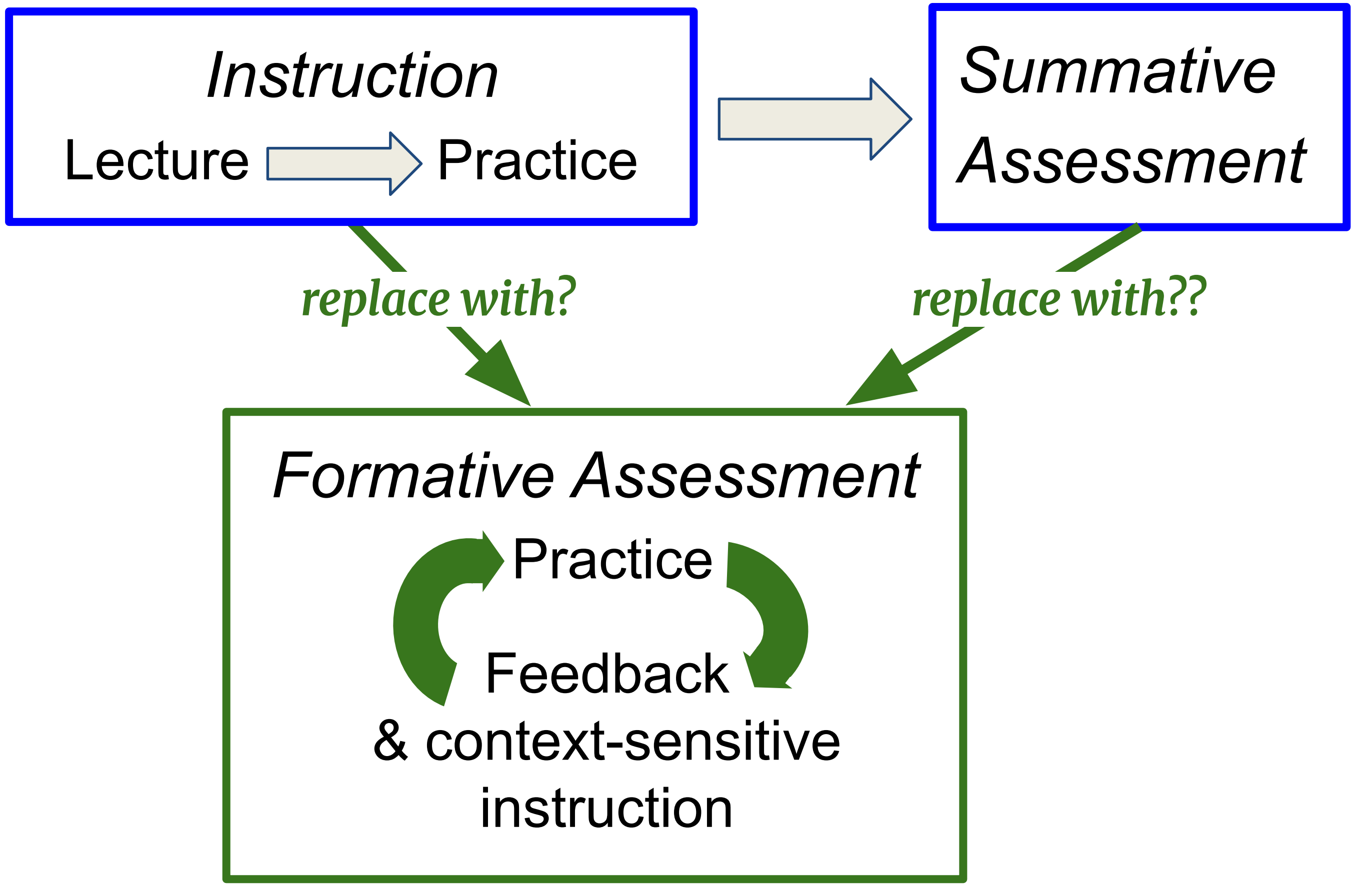
Can hybrid human-AI tutoring enhance educational equity?

Chine et al (2022). Educational Equity Through Combined Human-AI Personalization. AIEd Conference.

Holstein et al (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

How about summative assessment?

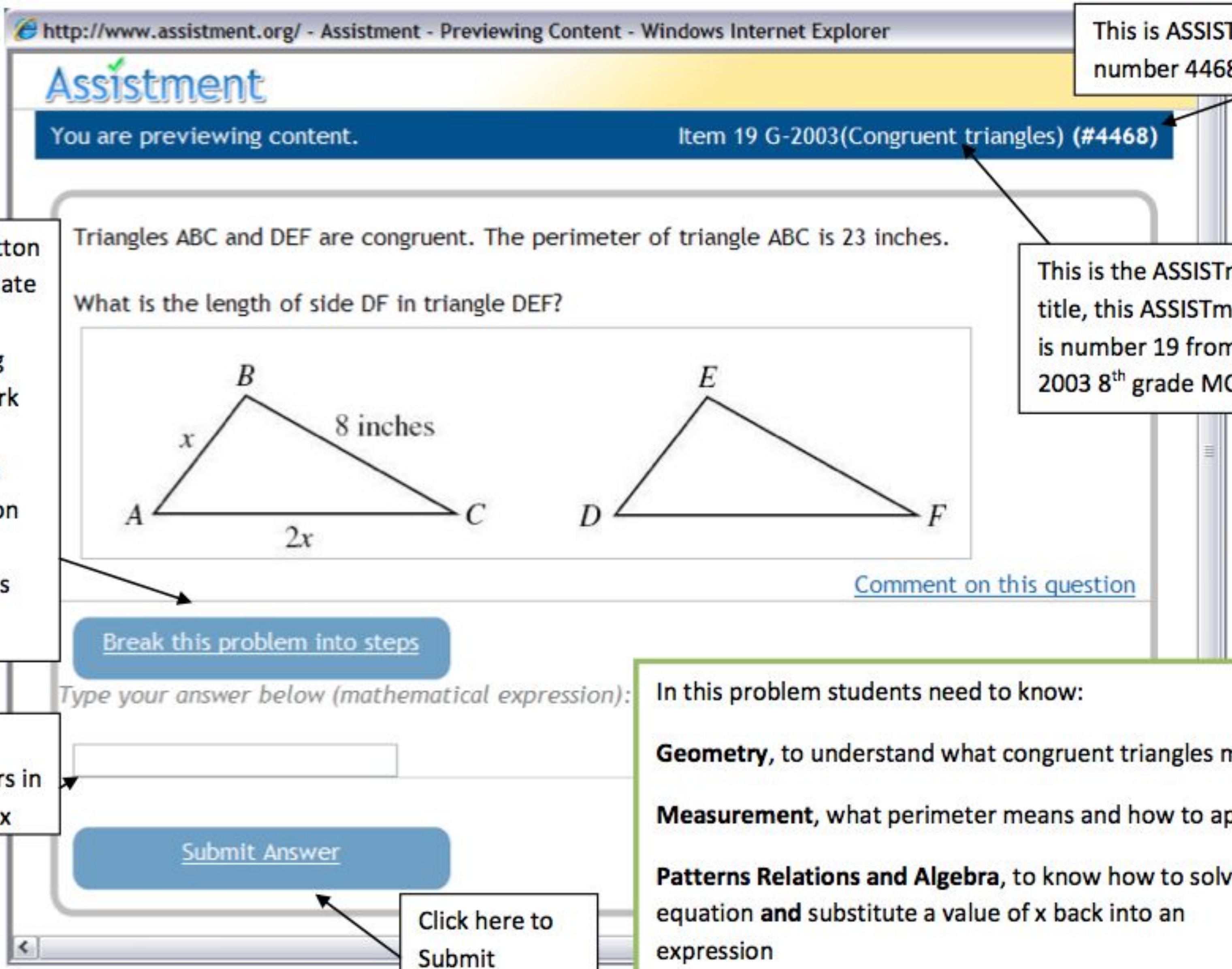
Most current education:



A vision for summative assessment achieved formatively

- Assessment is embedded in process of learning
- No waste of instructional time on tests
- Teachers, students, parents get fast & reliable info on student strengths & weaknesses

ASSISTment Practice Question



This is ASSISTment number 4468

This is the ASSISTment title, this ASSISTment is number 19 from the 2003 8th grade MCAS.

This button will initiate the tutoring and mark the student wrong on the teachers report.

Type answers in this box

Click here to Submit

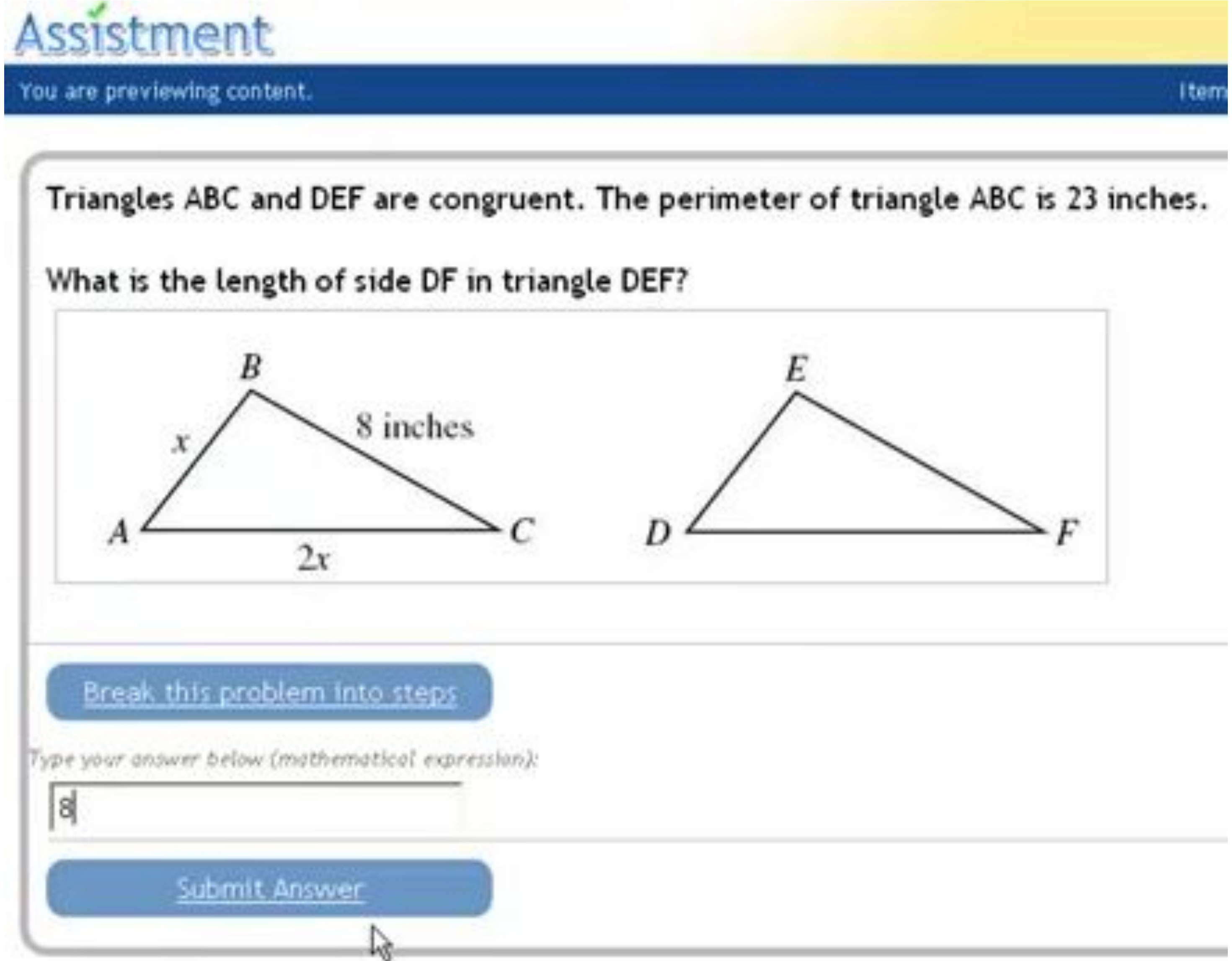
In this problem students need to know:

Geometry, to understand what congruent triangles mean

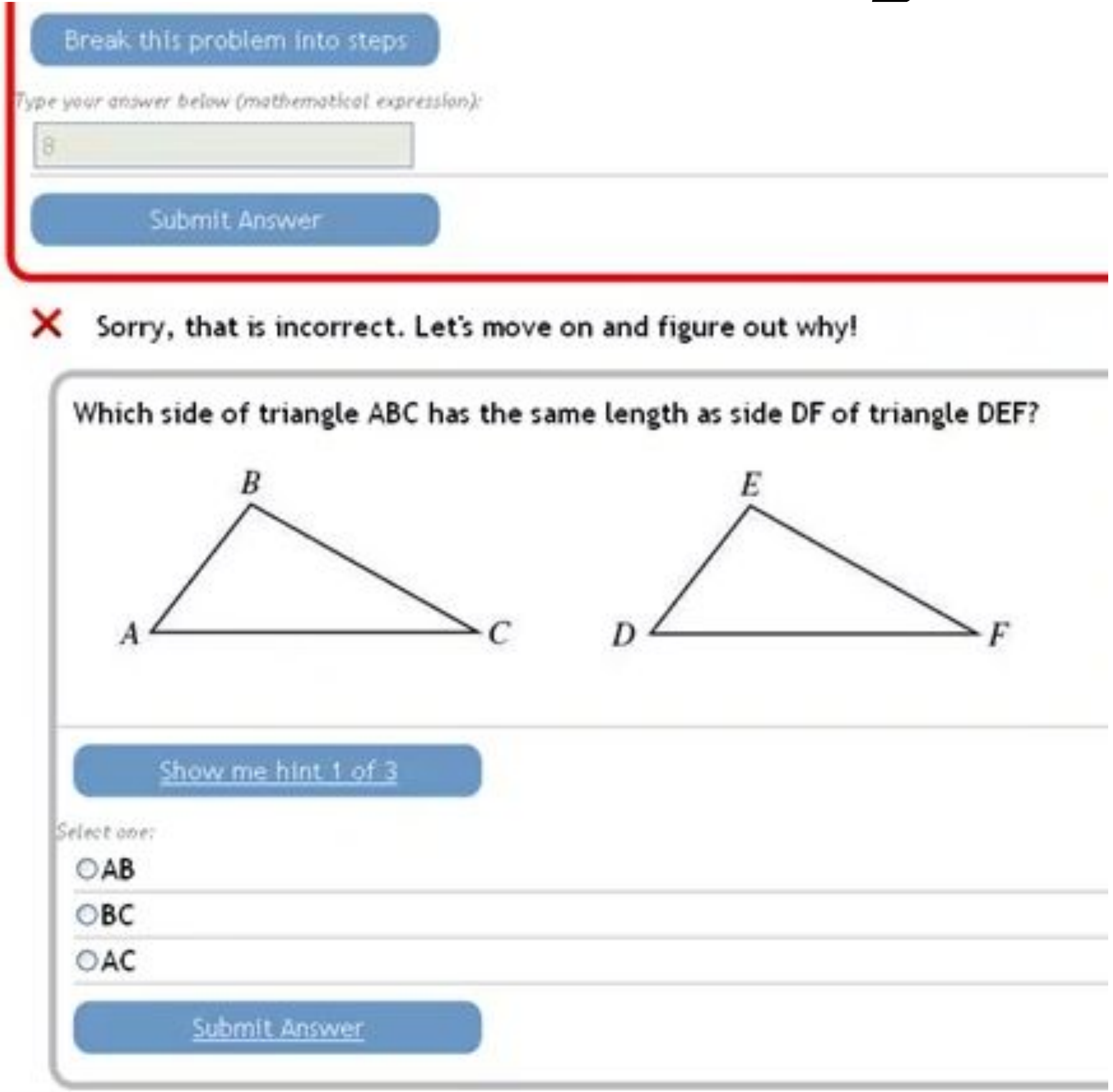
Measurement, what perimeter means and how to apply it

Patterns Relations and Algebra, to know how to solve an equation **and** substitute a value of x back into an expression

If answer to original question is wrong …

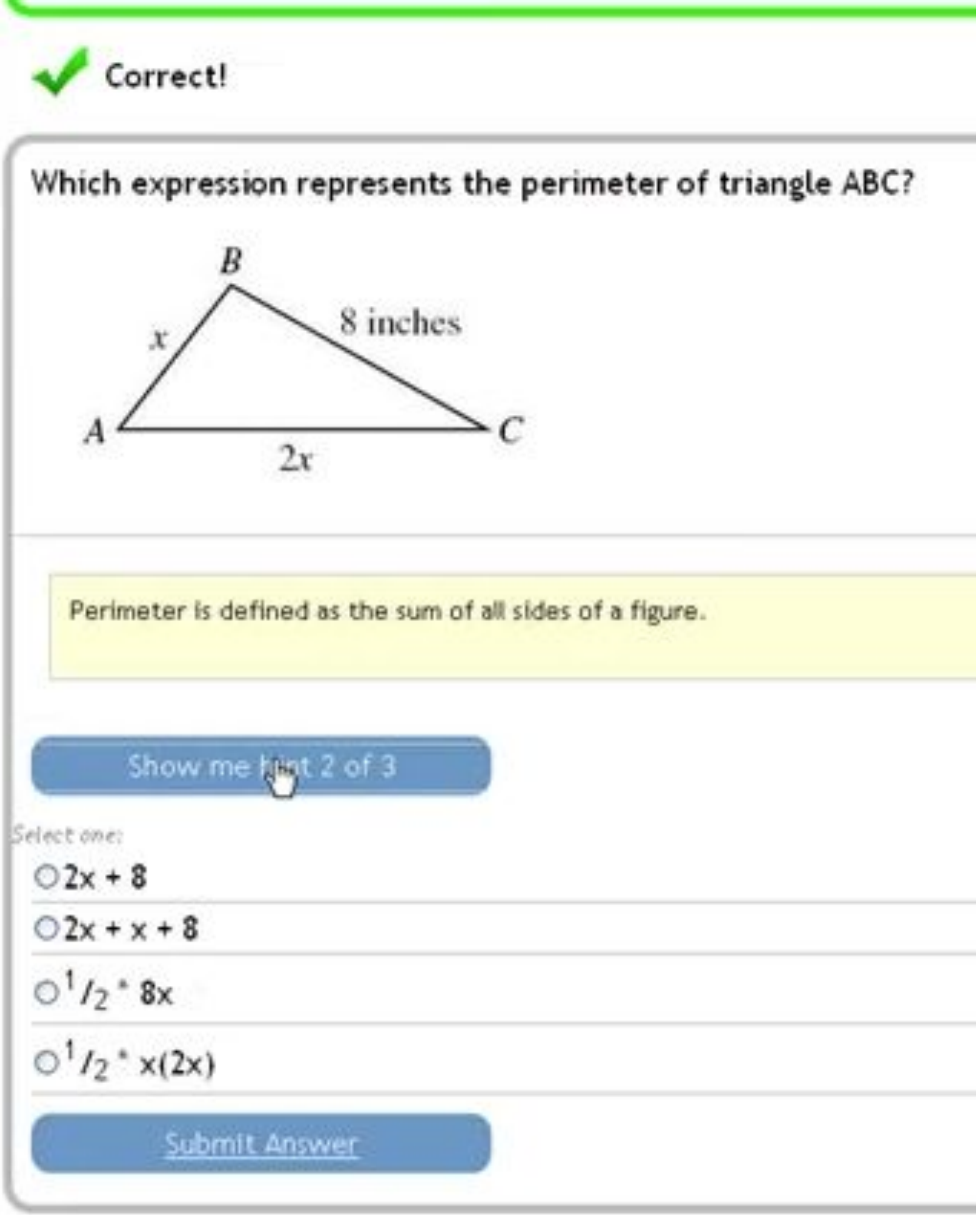


ASSISTment provides “scaffolding” questions that both *diagnosis* & *instruct*



MCAS tags as **geometry**.

Answers to scaffolds indicate **algebra** is key challenge



Does formative assessment accurately predict summative results?

Yes!
Models of online practice accurately predict end-of-year tests
- In ASSISTments
- In Cognitive Tutors

Ayers & Junker (2008). IRT Modeling of Tutor Performance To Predict End-of-year Exam Scores. *Educational and Psychological Measurement.*

Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*.

Ritter et al (2013). Predicting standardized test scores from cognitive tutor interactions. In Educational Data Mining.

Formative & Summative Data

- 2004-2005 Data, N = 391 students
  - Formative data: Sept - May
    - Means: 4:27 hours; 147 items
  - Summative: 8th grade MCAS state test in May

- 2005-2006 Data, N = 616 students
  - Formative data: Sept – May
    - Means: 3:16 hours; 88 items
  - Summative: 8th grade MCAS state test in May

Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*.

Results of regression

	2004-2005 data	2005-2006 data
(Constant)	32.414	3.284
IRT_Proficiency_Estimate	26.8	32.944
Scaffold_Percent_Correct	20.427	21.327
Avg_Question_Time	-0.17	-0.102
Avg_Attempt	-10.5	
Avg_Hint_Request	-3.217	
Question_Count		0.072
Avg_Item_Time		0.045
Total_Attempt		-0.044
Correlation with MCAS	**R = .84**	**R = .85**

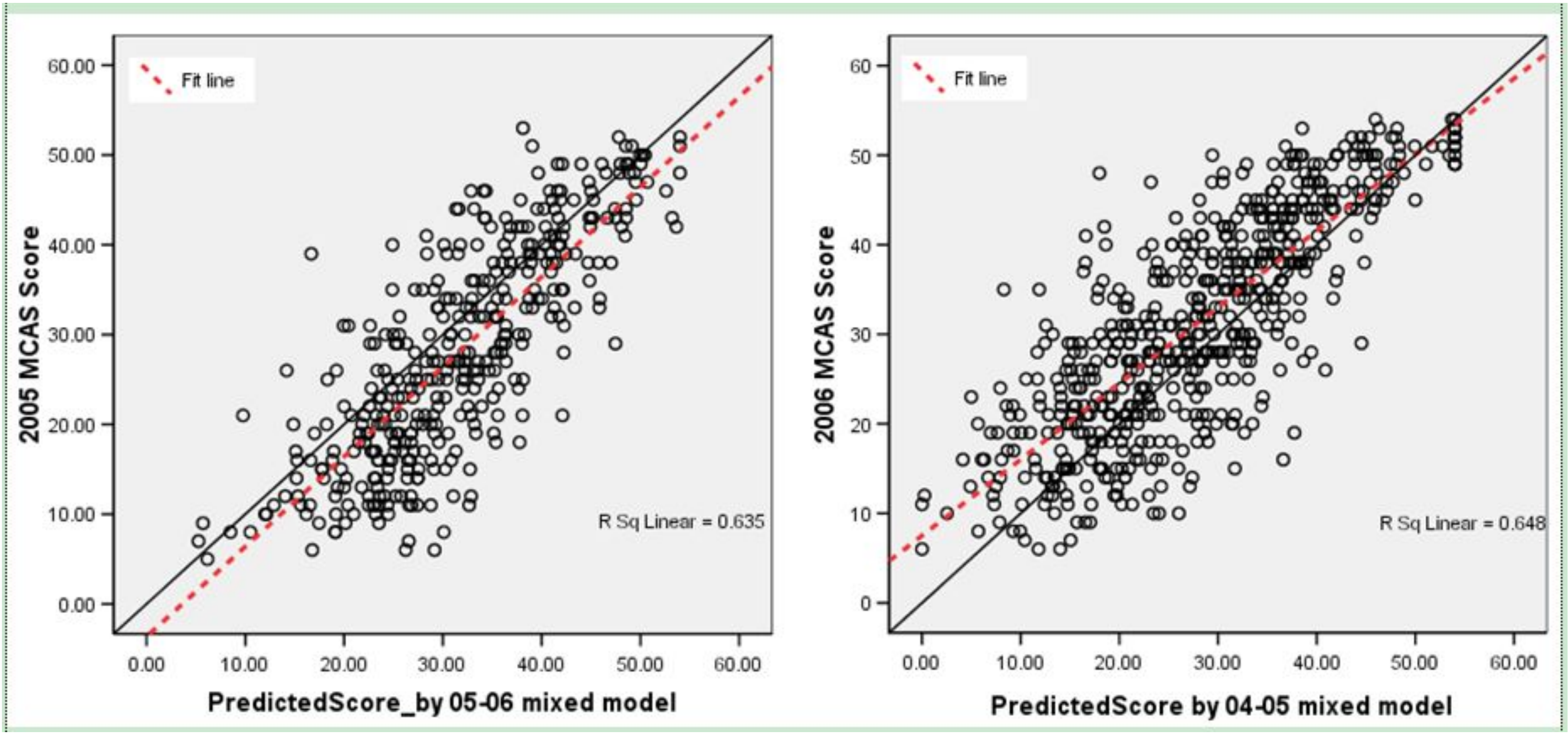
Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research.*

Models generalize

	2004-2005 data	2005-2006 data
(Constant)	32.414	3.284
IRT_Proficiency_Estimate	26.8	32.944
Scaffold_Percent_Correct	20.427	21.327
Avg_Question_Time	-0.17	-0.102
Avg_Attempt	-10.5	
Avg_Hint_Request	-3.217	
Question_Count		0.072
Avg_Item_Time		0.045
Total_Attempt		-0.044
Correlation with MCAS	**R = .84**	**R = .85**
Other-year MCAS corr	***R = .83***	***R = .82***

Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*.

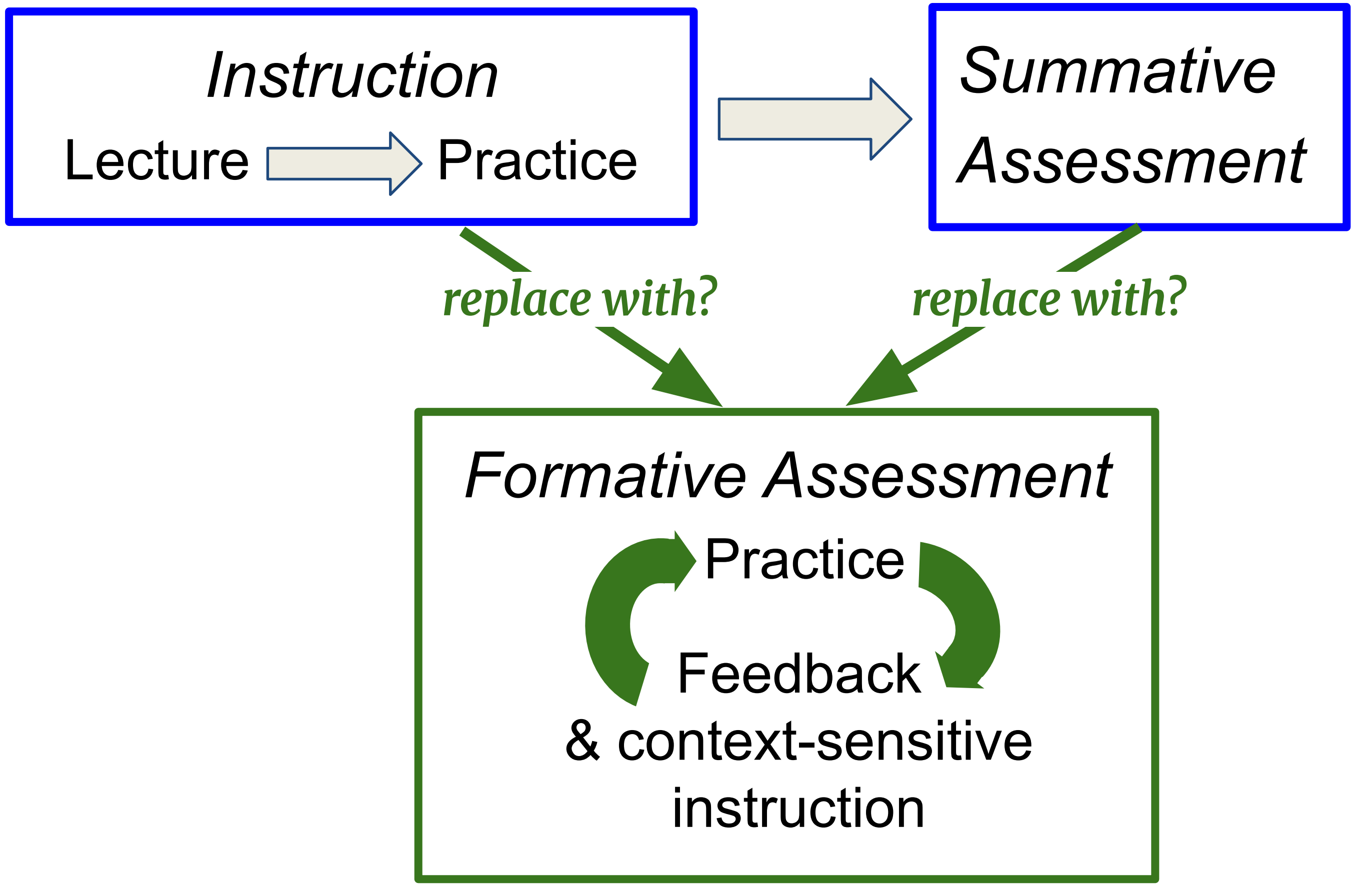
Cross-year validation: Prediction models are robust!



Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*.

Formative assessment accurately predicts summat
assessment

Most current education:



Overview

Interactive learning by doing & online tutors

What student differences account for learning outcome differences?

Koedinger, Carvalho, Liu, McLaughlin (2023). An Astonishing Regularity in Student Learning Rate. *PNAS*.

Formative Assessment instead of Summative Assessment

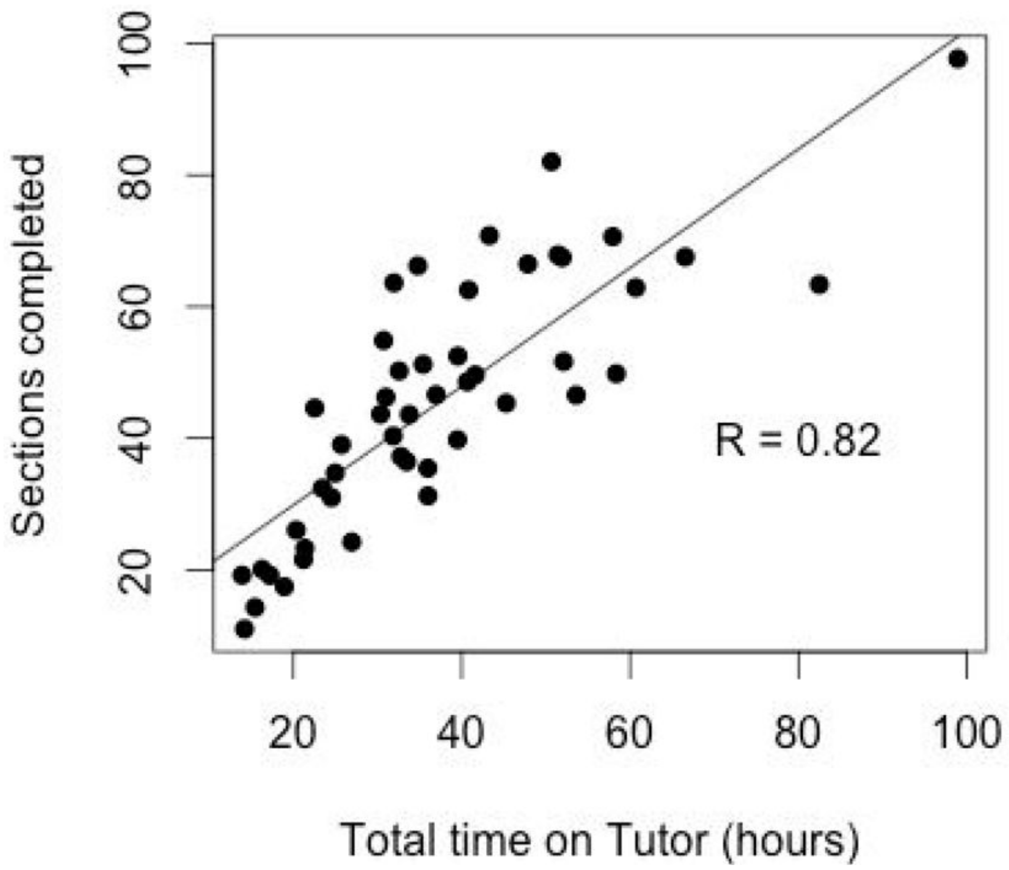
Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*.

Can hybrid human-AI tutoring enhance educational equity?

Chine et al (2022). Educational Equity Through Combined Human-AI Personalization. AIEd Conference.

Holstein et al (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

Deliberate practice works …
for those who engage in it

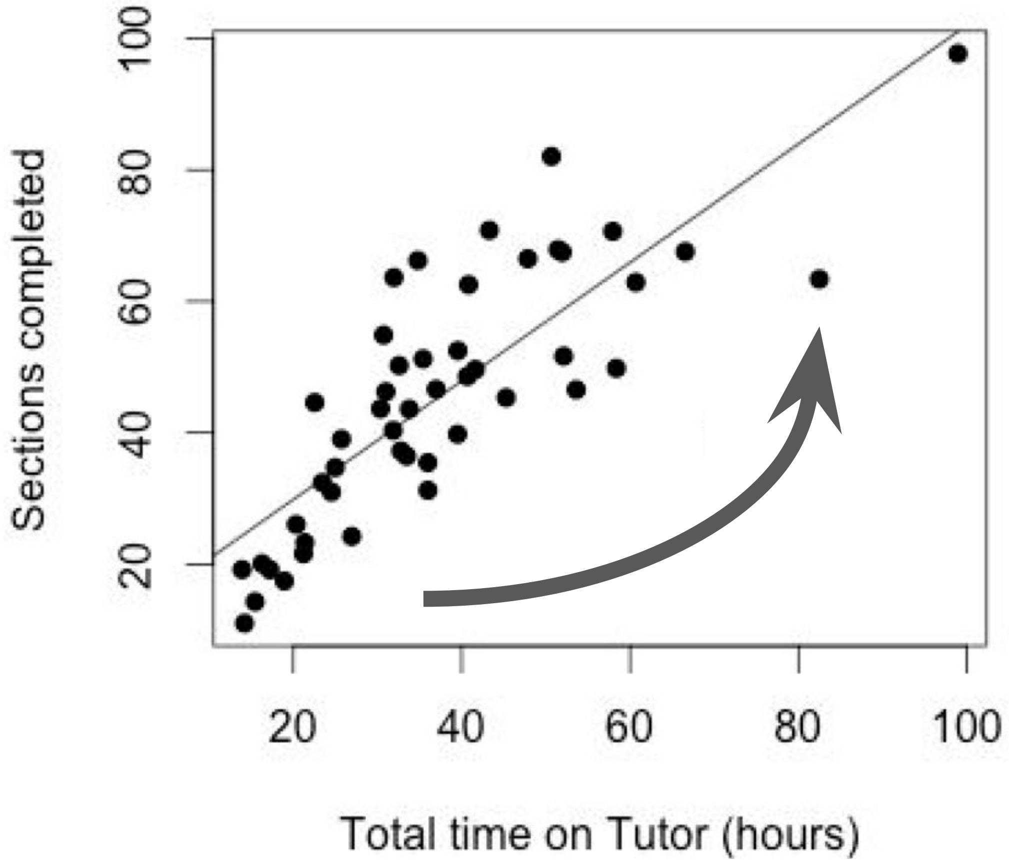


Data from online tutors:

Opportunity gaps

yield achievement gaps

How to address opportunity gaps?



Many options from social psychology motivation theory

Growth mindset

Belonging

Utility-value

Relationship building

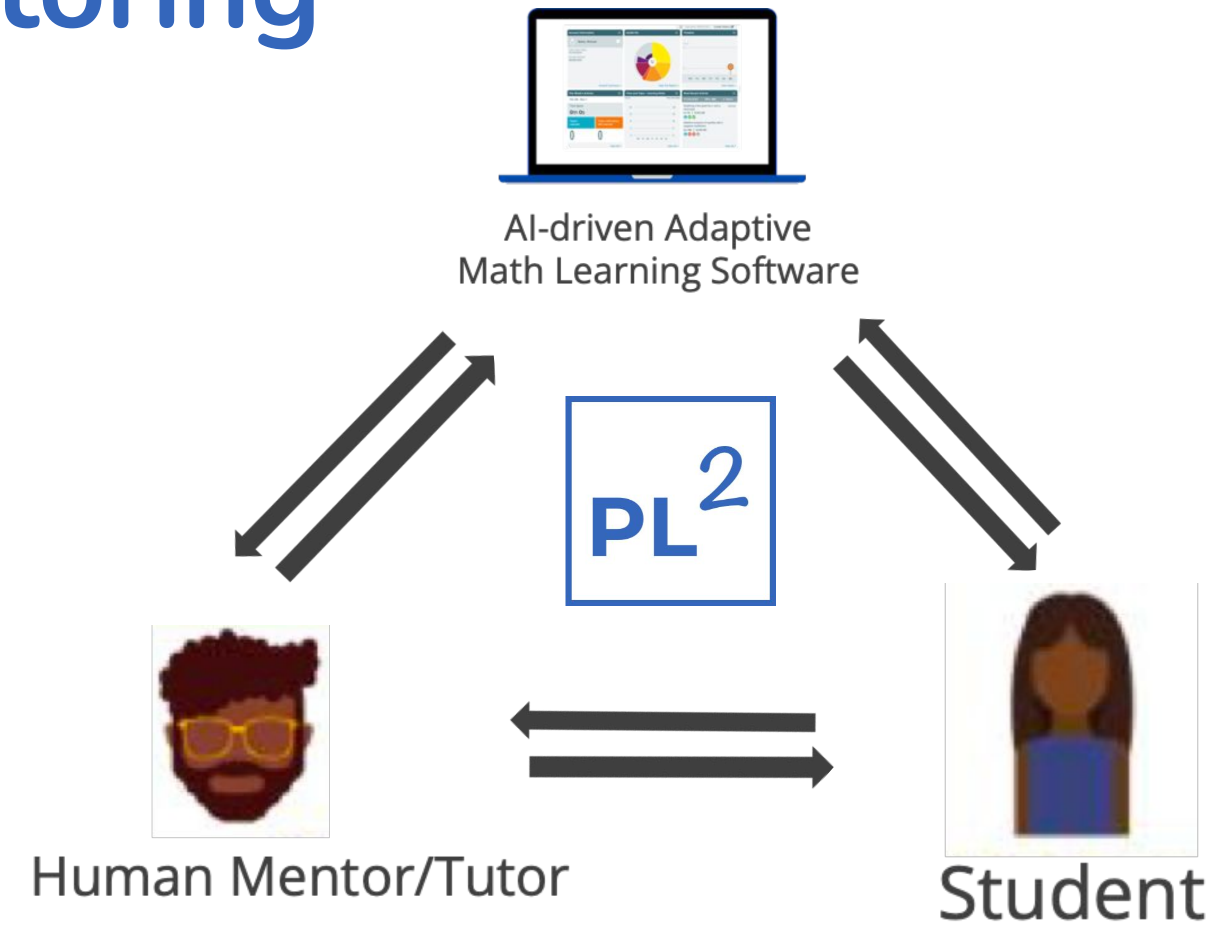
Stereotype threat

Personalized Learning Squared (PLUS): Hybrid human-AI tutoring

Complementary strengths

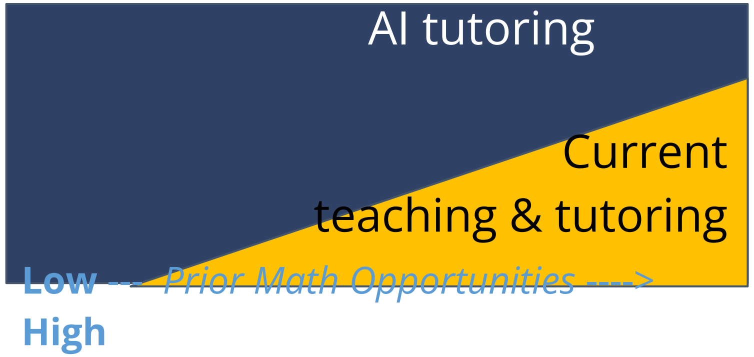
- AI tutors support math learning
- Human tutors support motivation

Data from AI tutors guides human

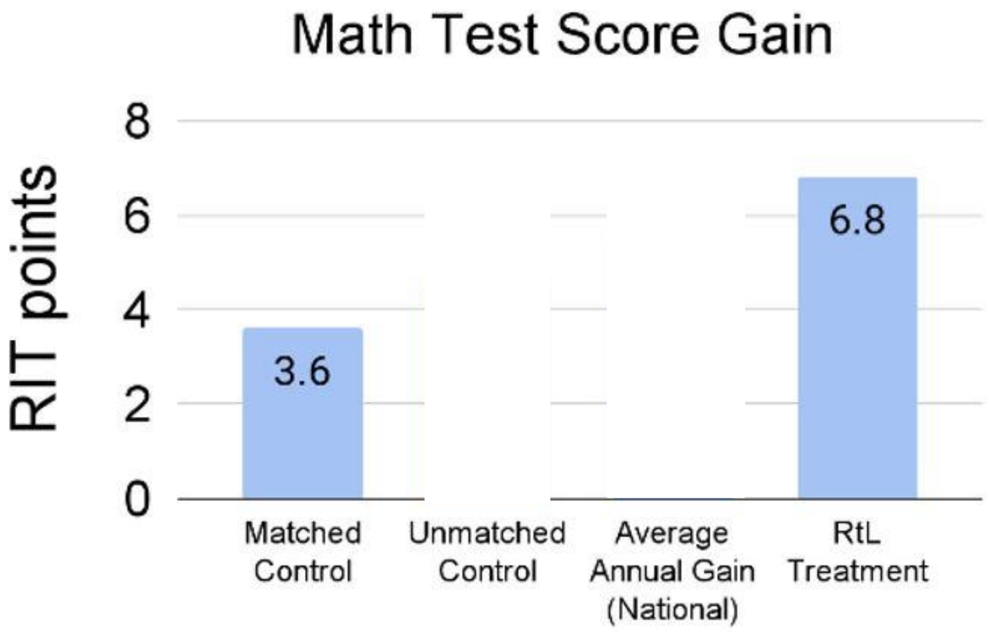


Goal: Reallocate human attention to achieve equity

Strategy	Who gets more attention?
Current practices	**Higher** prior opportunity students
Dashboard-driven tutoring	**Lower** prior opportunity students



PLUS afterschool tutoring:
Nearly *doubled* math learning during pandemic



Significant Treatment x PostTest interaction (p<.01)
=>
Greater learning for treatment

Effect size: .4 sd

Math Score Difference Results

Variable	***Estimate***	***SE***	***df***	***t***	***p***
(Intercept)	211.16	0.79	520.30	267.17	**0.000*****
Treatment	3.82	2.00	520.30	1.90	0.057
PostTest	3.60	0.43	448.00	8.30	**0.000*****
Treatment x PostTest	3.21	1.10	448.00	2.92	**0.004****

Chine et al (2022). Educational Equity Through Combined Human-AI Personalization: A Propensity Matching Evaluation. AIEd Conference.

Two paths to reach more kids in school

1. Teachers as tutors
2. Undergrads as remote tutors zooming into class

Both guided by data from AI tutors

Smart Glasses Dashboard for Teachers as Tutors

Teacher wears smart glasses

Sees struggle icons over students in class

“Misusing” the software
(e.g., rapid guessing, abusing hints, gaming-the-system)

“Unproductively” struggling
(e.g., many attempts but low mastery, high frustration)

Struggling
(e.g., lots of errors, but not necessarily *unproductive struggle*)

Recently doing “very well”

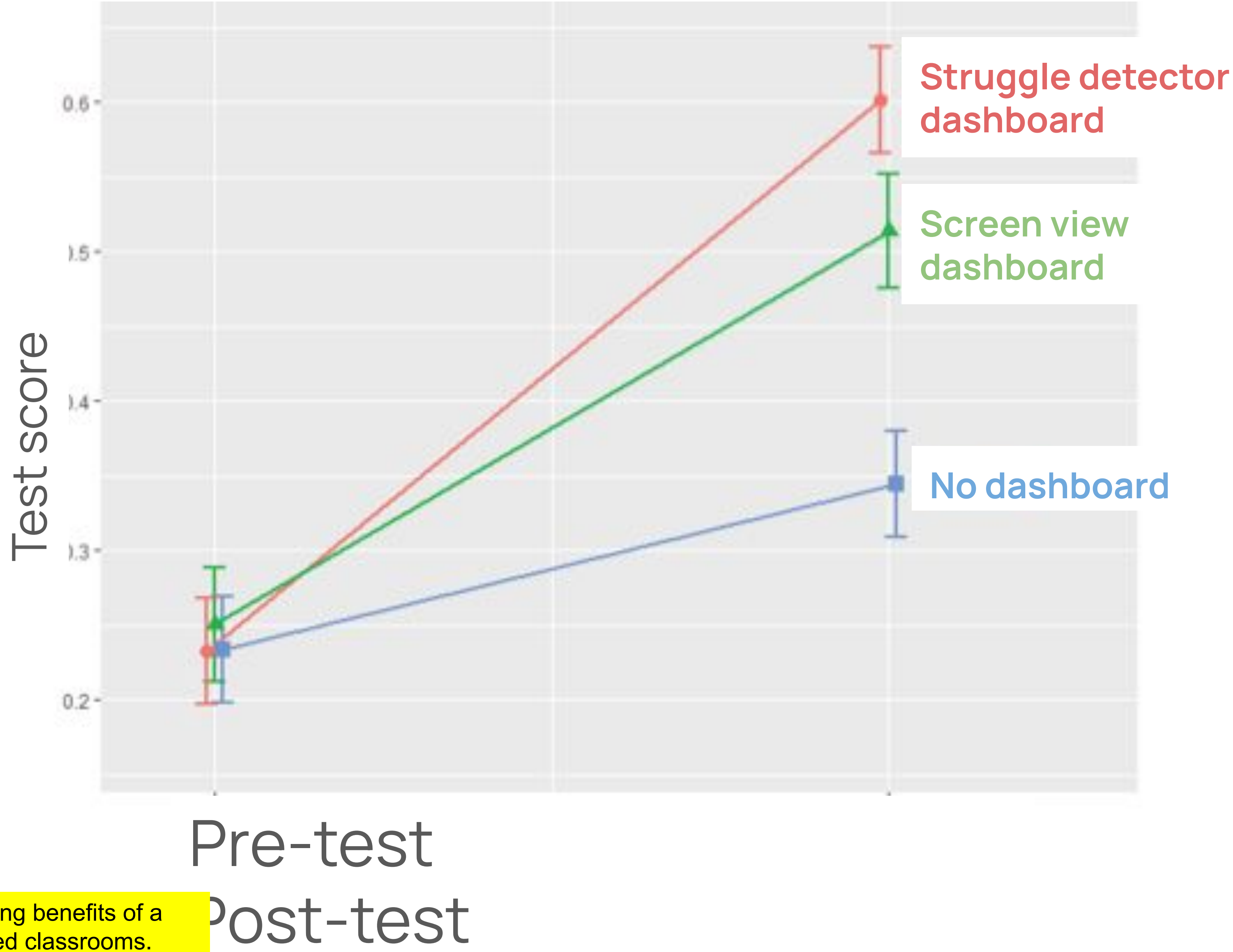
Idle

Holstein, McLaren, & Aleven (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

Shifts teacher attention to low prior opportunity students

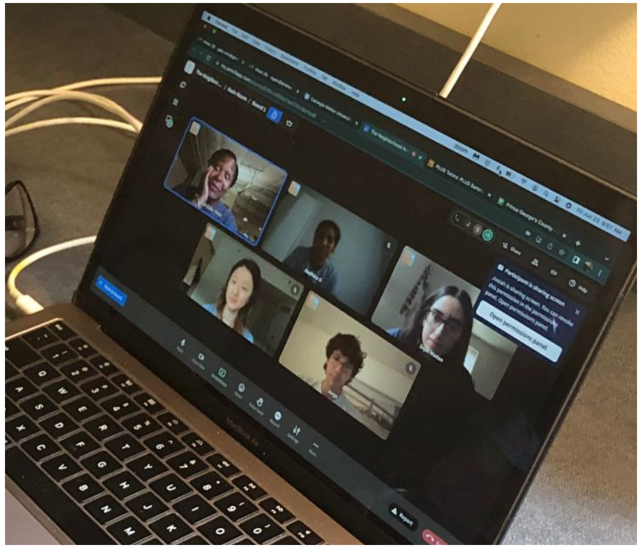
Holstein, McLaren, & Aleven (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

Students learn 3x more math!



Holstein, McLaren, & Aleven (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

PLUS: Remote Tutors use AI Dashboard & zoom in to help



AI Struggle Detectors

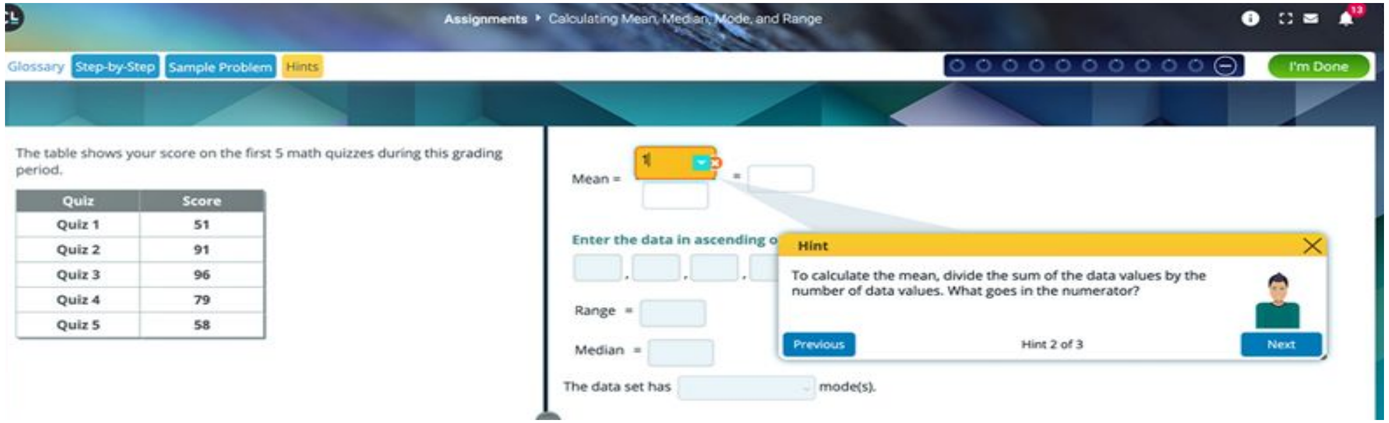
- Slow start
- Idle
- Gaming the system
- Unproductive struggle
- Productive struggle
- Doing well

Dashboard to guide tutors to students with greatest needs

My Students View All

Name	Status	Focus Area
Hermione Granger	Ramp it up	Mastering Content
Ron Weasley	Missed you	Social-Emotional
Harry Potter	Wow	Advocacy

Math Ed Tech



Summary

Online interactive learning by doing provides favorable learning conditions

Lectures/readings => 65%

7 deliberate practice opportunities => 80%

Students in same course:

differ widely in initial knowledge – 55% to 75%

astonishingly similar in learning rate – 1.7% to 2.6% per opp

Koedinger, Carvalho, Liu, McLaughlin (2023). An Astonishing Regularity in Student Learning Rate. *PNAS*.

Formative Assessment: replace lecture & summative assessment?

Feng, Heffernan, & Koedinger (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*.

Hybrid human-AI tutoring shows promise for educational equity

Chine et al (2022). Educational Equity Through Combined Human-AI Personalization. AIEd Conference.

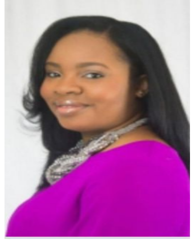
Holstein et al (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. AIEd Conference.

Thank you!

Ken Koedinger with co-authors



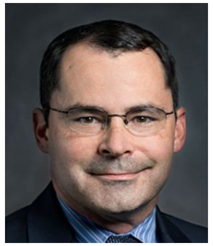




Paulo Carvalho Mimi McLaughlin Norman Bier Ran Liu Cassandra Brentley



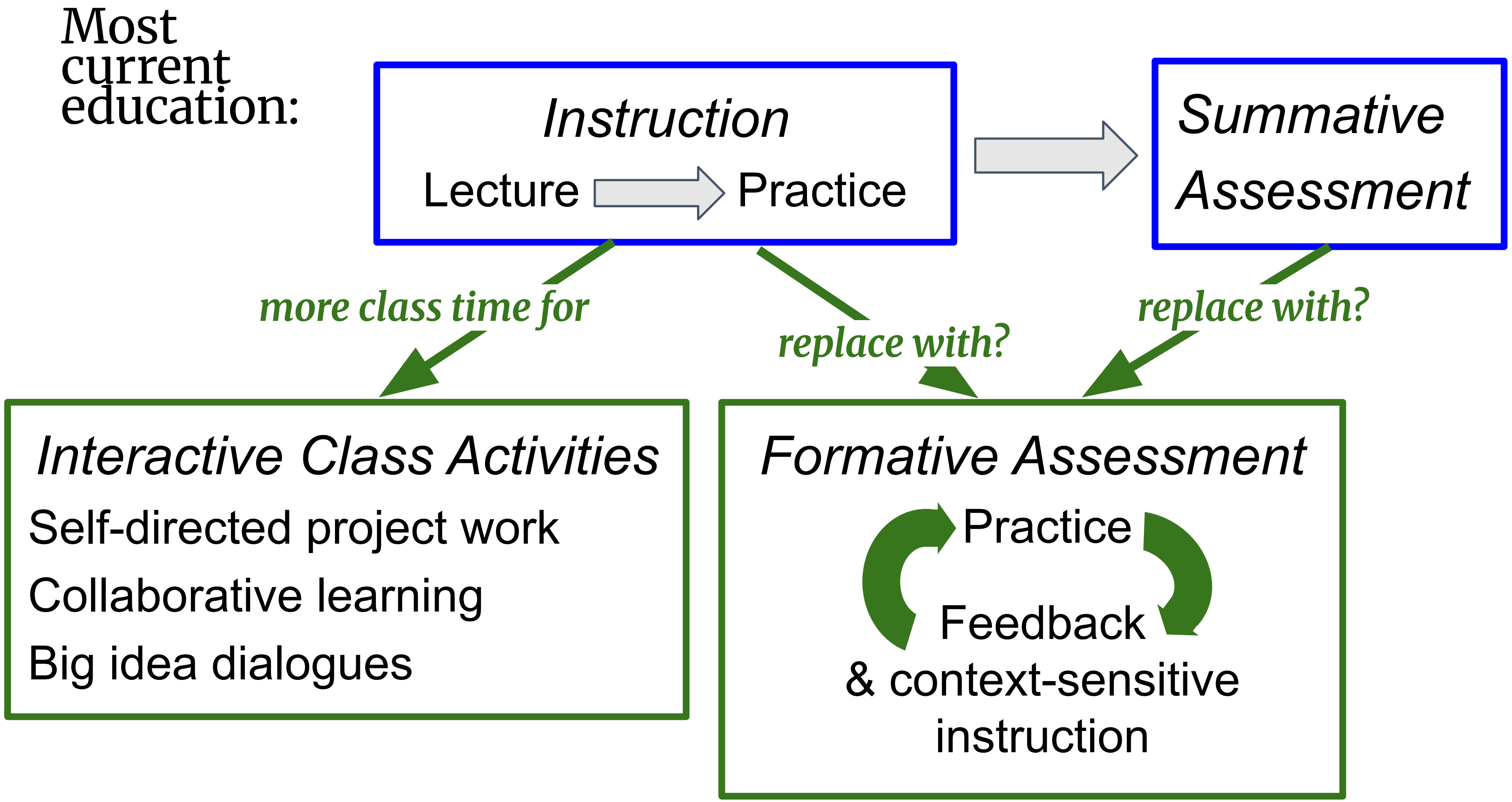




Danielle Chine/Thomas Liz Richey Menaf Gul Carmen Thomas-Browne Lee Branstetter

Thanks to **NSF** for funding of LearnLab & LearnSphere & to **CZI, RK Mellon, Schmidt** for funding of PLUS

Should we totally rethink education?



EXTRAS

Theory: Why is individual learning rate highly regular?

Koedinger, Corbett, & Perfetti (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*.

Disjunctive learning paths hypothesis v1

Explanation-based learning is flexible

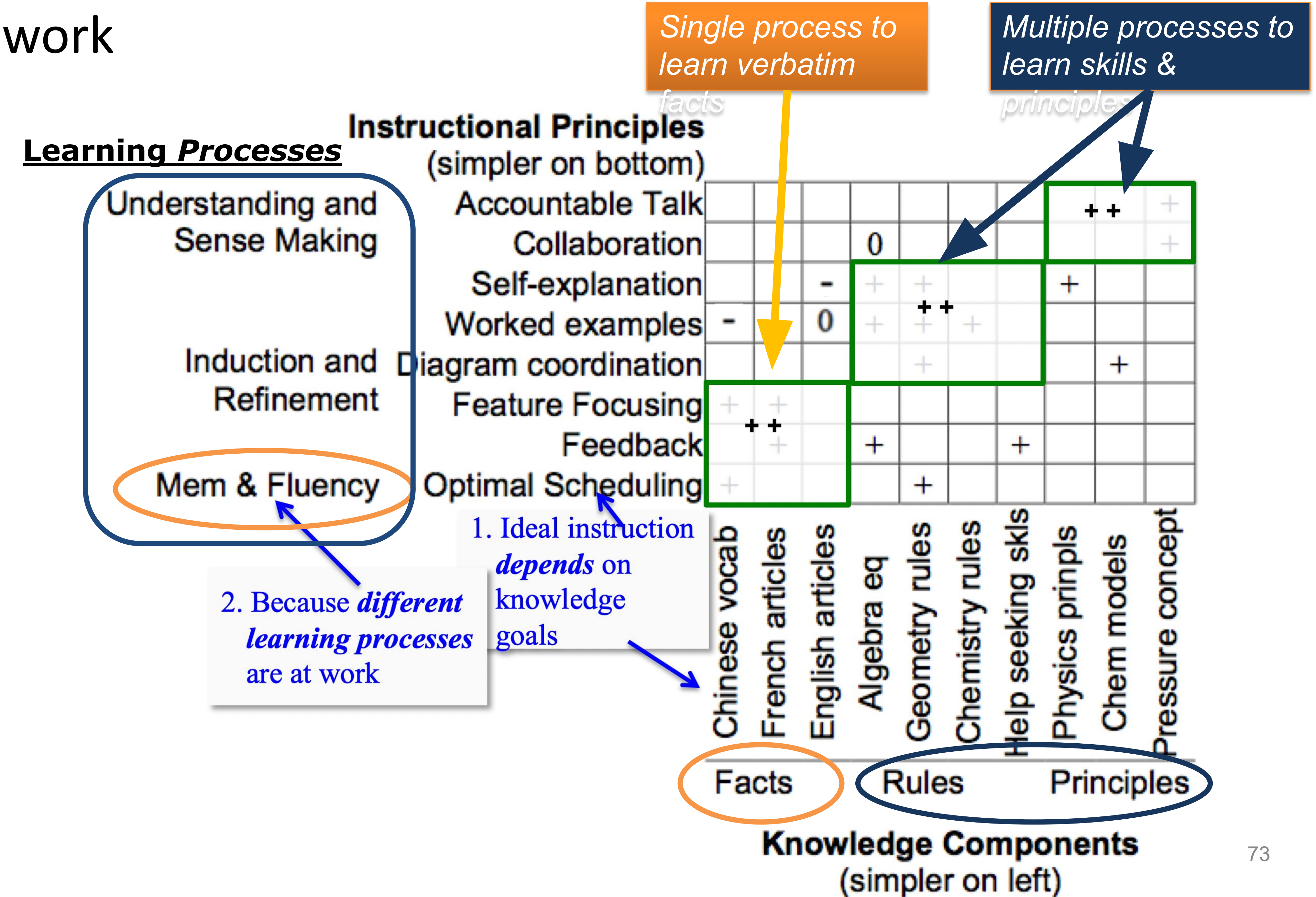
Maclellan et l (2016). The Apprentice Learner Architecture: Closing the Loop between Learning Theory and Educational Data. *Int Ed Data Mining Society*.

- AI-based learning theory (AL):
  - Implicit, inductive self-explanations of examples based on background k
  - Different self-explanations work in AL

  => Learners are not dependent on a single set of prerequisites

- Given an algebra solution example (2x - 5x = -3x)
  - S1 self-explains using symbolic negative number knowledge
  - S2 self-explains using a number line

KLI Framework



Disjunctive learning path hypothesis v2

In many domains, especially math & science

- Content: generalizable skills & re-discoverable principles
- Students can learn by any of:
  verbatim memory ***OR*** *pattern induction* ***OR*** *sense-making*

=> Less optimal learning of one kind can be compensated by another

In some domains, such as language

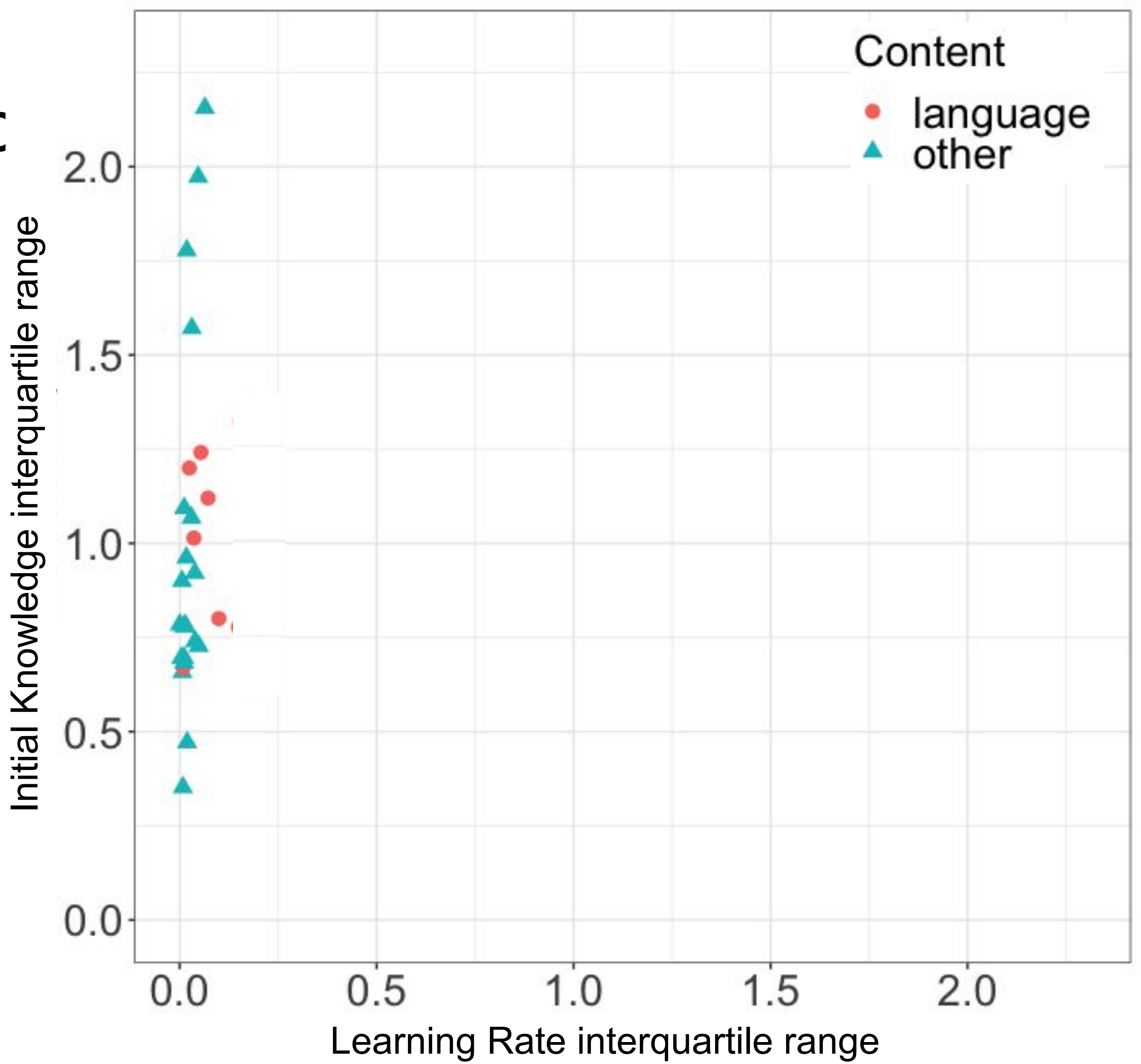
- Content: arbitrary mappings (e.g., the Pacific Ocean; ~~the~~ Lake Michigan)
- Students must learn mostly by:
  verbatim memory

=> Core verbatim/episodic memory limitations produce bigger differences

Visualizing variability in each c 27 datasets

Student
initial knowledge is
10x more variable than
learning rate

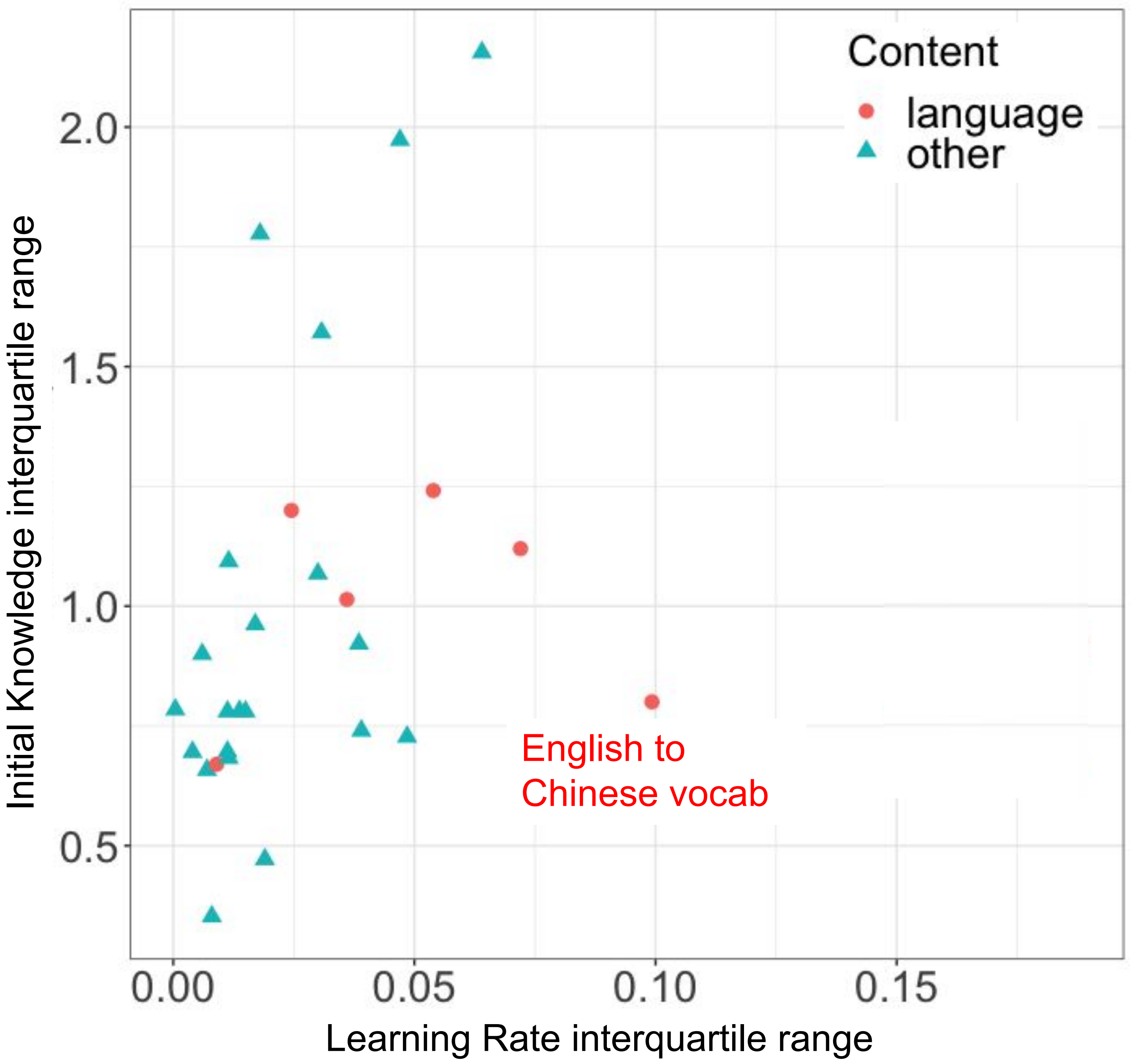
Variability measure = Interquartile range of parameter estimates in log odds



Stretching out learning rate by ~10x

More rate variation in language learning

Esp for verbatim recall of paired-associates



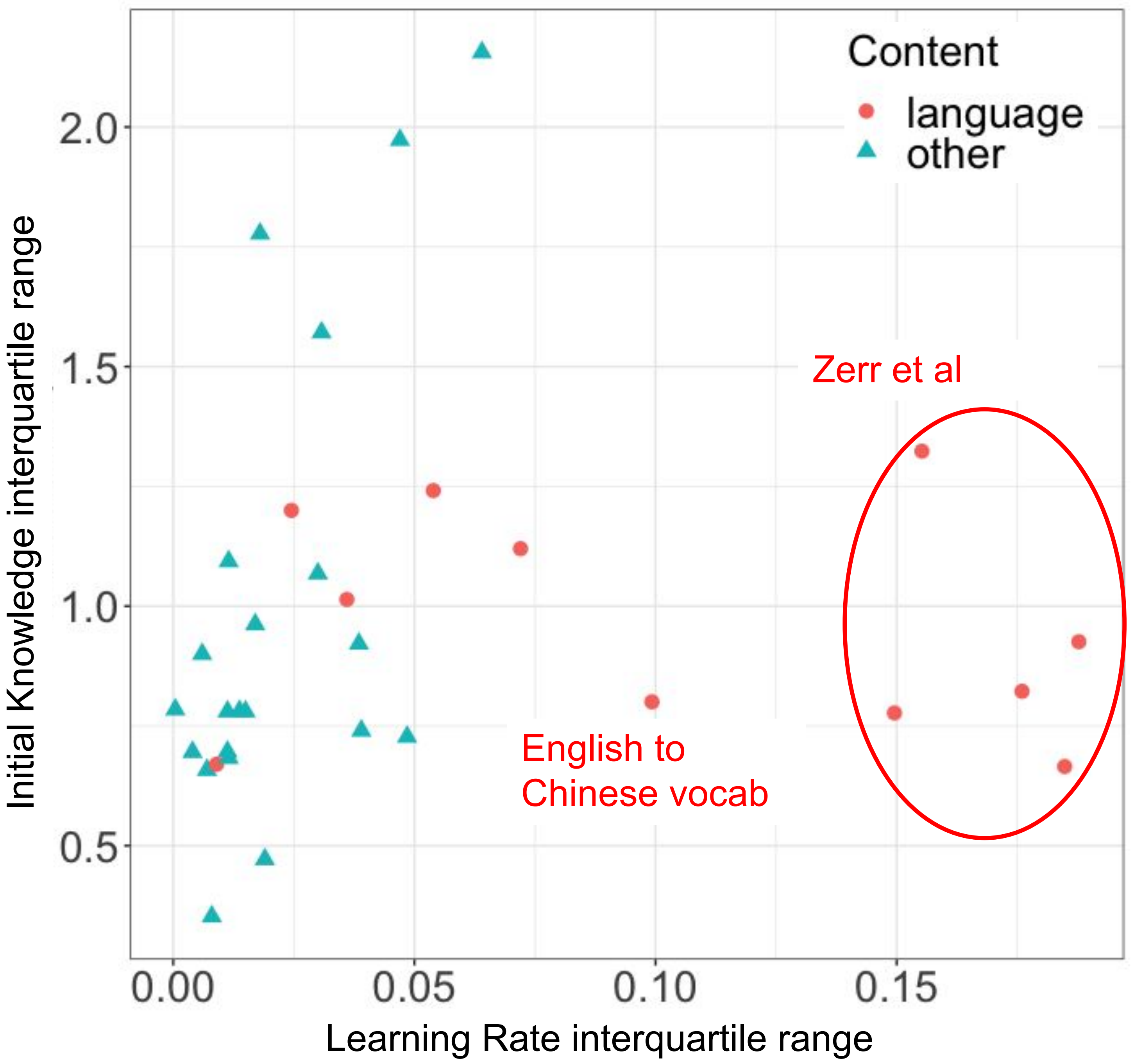
Other verbatim memory learning rate estimates

Zerr et al data

- Lithuanian (KNYGA – BOOK) Chinese (风 – WIND), & other paired-associate datasets

Also higher variation!

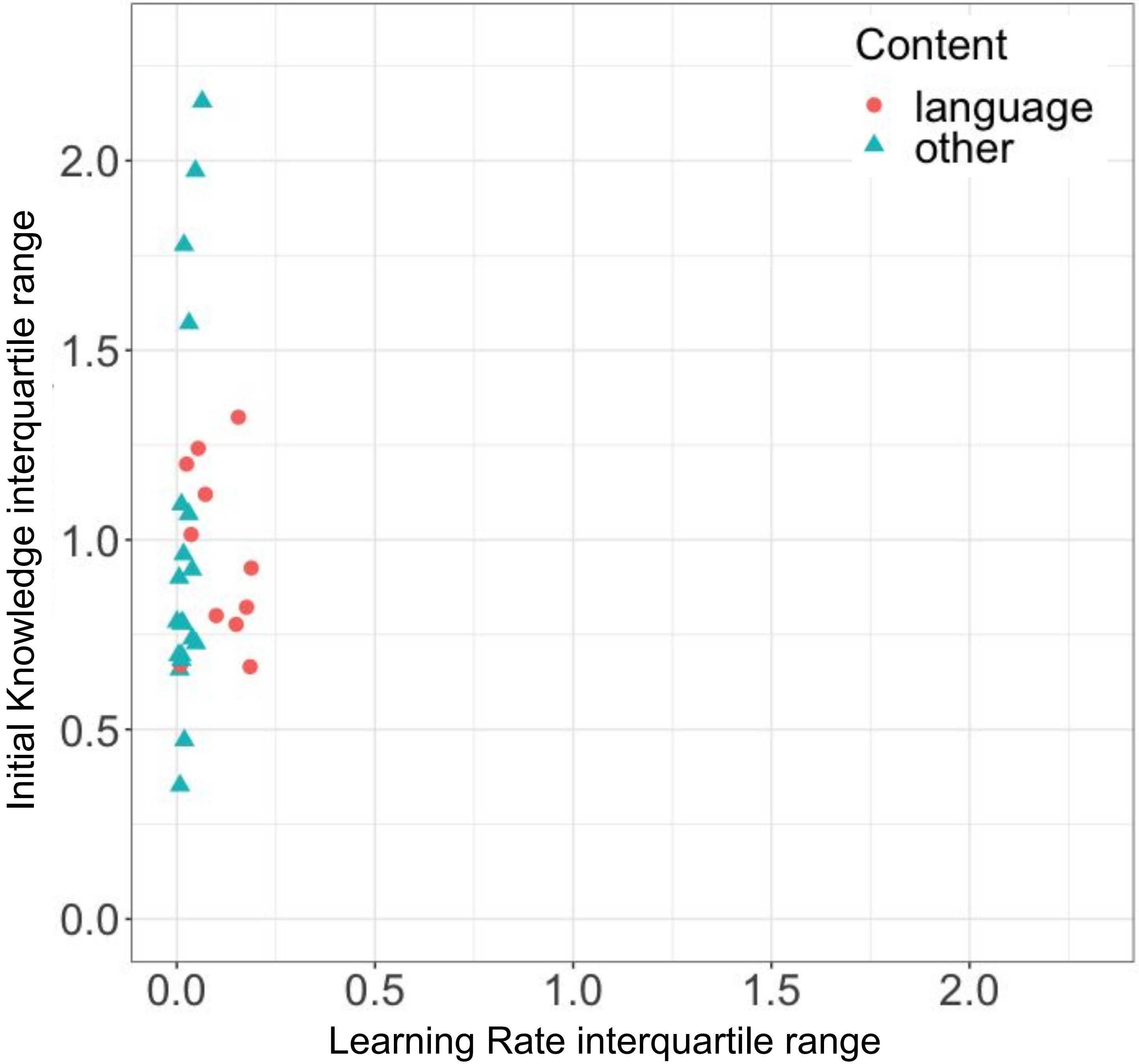
- (Some extra variation comes from difference in method)



Zerr et al. (2018) Learning efficiency: Identifying individual differences in learning rate and retention in healthy adults. Psych Sci.

Remember:

These learning rate variations are relatively quite small



Question for you!

Psychology is immensely relevant to our everyday lives at work, school, & home

Strangely, it does not seem so appreciated

I wonder why? Why not taught in schools?

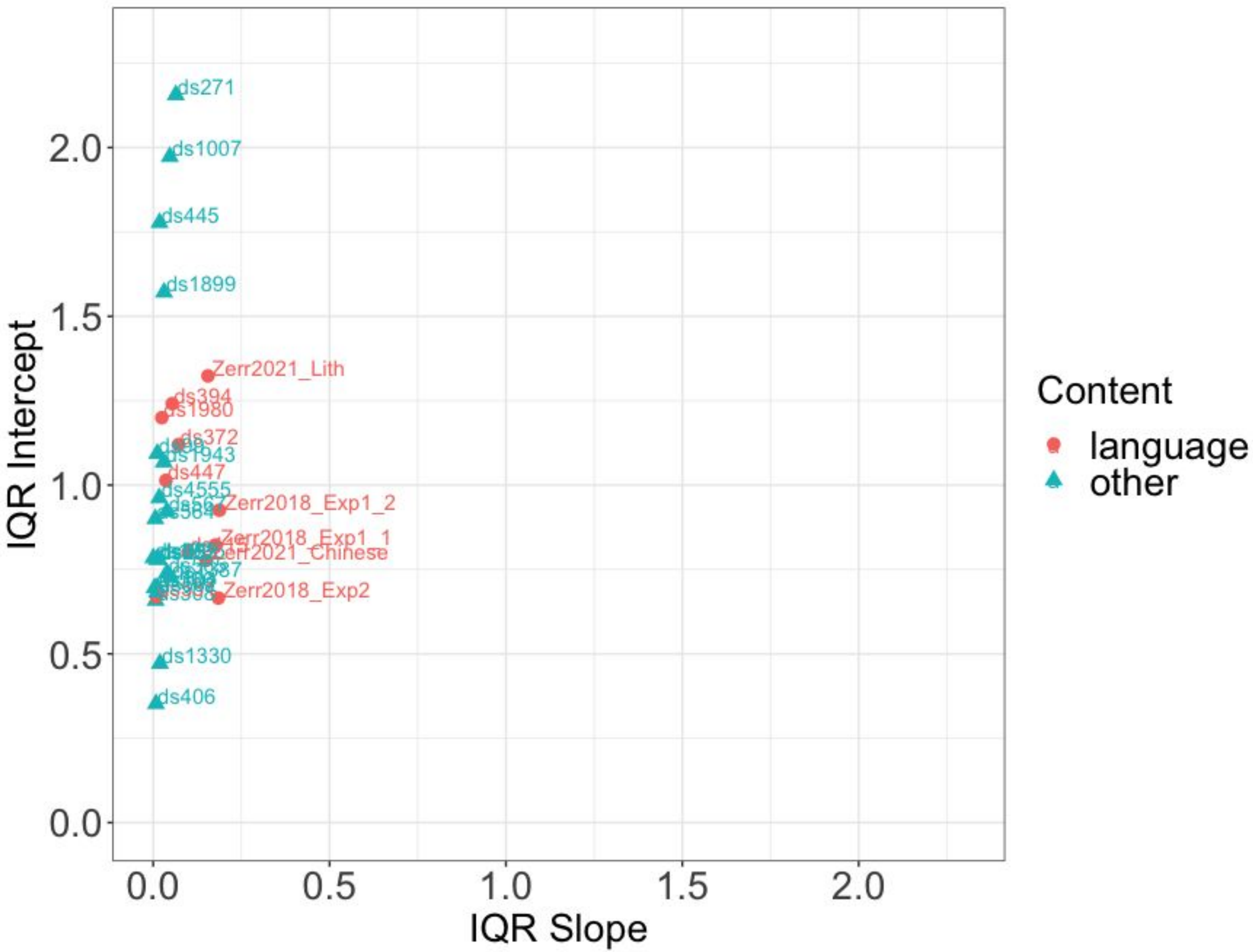
Is it not as important as plate techtonics?

memory differences & memory-induction trade-off

- https://www.jstor.org/stable/pdf/2885095.pdf
- Evidence of stable individual differences in implicit learning
  - "low variability across individuals was an important part of Reber's initial hypothesis that implicit learning differed fundamentally from the well-known individual variability in explicit learning."
  - See reber paper
- Category learning and the memory systems debate - Poldrack

Abstract. I will talk about our recent 2023 PNAS paper with the following abstract: Leveraging a scientific infrastructure for exploring how students learn, we have developed cognitive and statistical models of skill acquisition and used them to understand fundamental similarities and differences across learners. Our primary question was why do some students learn faster than others? Or, do they? We model data from student performance on groups of tasks that assess the same skill component and that provide follow-up instruction on student errors. Our models estimate, for both students and skills, initial correctness and learning rate, that is, the increase in correctness after each practice opportunity. We applied our models to 1.3 million observations across 27 datasets of student interactions with online practice systems in the context of elementary to college courses in math, science, and language. Despite the availability of up-front verbal instruction, like lectures and readings, students demonstrate modest initial pre-practice performance, at about 65% accuracy. Despite being in the same course, students' initial performance varies substantially from about 55% correct for those in the lower half to 75% for those in the upper half. In contrast, and much to our surprise, we found students to be astonishingly similar in estimated learning rate, typically increasing by about 0.1 log odds or 2.5% in accuracy per opportunity. These findings pose a challenge for theories of learning to explain the odd combination of large variation in student initial performance and striking regularity in student learning rate.

In addition to discussing these recent analyses, I will describe recent efforts pursuing the hope inherent in this evidence: That given favorable learning conditions for deliberate practice and given the learner invests effort in sufficient learning opportunities, indeed, anyone can learn anything they want. In particular, we have been experimenting with cost-effective methods to provide math students with extra human tutoring toward increasing their motivation to engage in practice and we have demonstrated promise in reducing achievement gaps by so reducing opportunity gaps.


IQR Intercept
IQR Slope
0.0
0.5
1.0
1.5
2.0
ds271
ds1007
ds445
ds1899
Zerr2021_Lith
ds394
ds1980
ds372
ds1943
ds447
ds4555
Zerr2018_Exp1_2
Zerr2018_Exp1_1
Zerr2021_Chinese
Zerr2018_Exp2
ds1330
ds406
Content
language
other

Since writing the paper ...
There’s an issue with the disjunctive learning path hypothesis

KLI Framework suggests math & science need all 3 kinds of learning processes – AND not OR

In many domains, especially math & science

- Content: generalizable skills & re-discoverable principles
- Students ~~can learn by any of~~ need all of:
  verbatim memory ~~***OR***~~ ***AND*** *pattern induction* ~~***OR***~~ ***AND*** *sense-making*

Those better in verbatim memory should also learn math & science better …

but they don’t …

Learning process trade-off hypothesis

Individuals better at verbatim memory are worse at generalization

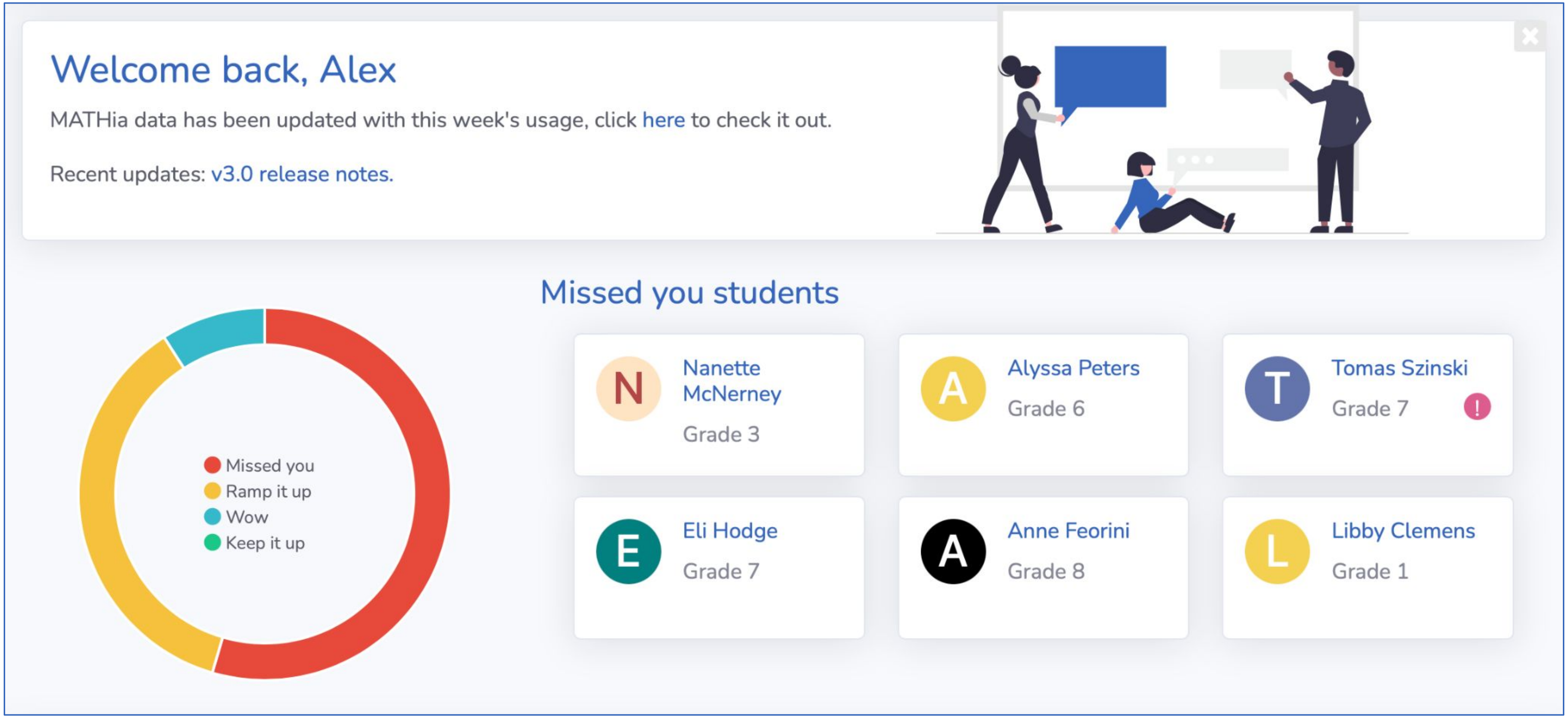
- Bias toward concreteness, high dimensionality

Individuals better at generalization are worse at verbatim memory

- Bias toward abstraction, low dimensionality

Prior evidence for such a trade-off?

PLUS is a Fitbit for learning
provides tutors with student effort & progress



PLUS tutors tutors!

Home > Resources for Mentors

Resource Assistant | View Pinned Resources | Create Resource

Building Mentor Understanding
Build your confidence by learning about math content standards, pedagogical practices, and how students learn

Building Relationships
Learn how to build relationships with students and parents

Maintaining Effort
Identify ways to support students' motivation, engagement, and goal-setting

Managing Emotions
Help students make more responsible decisions by building their emotional awareness and regulation

Navigating Technology
Get support for navigating online learning and using technology, including specific math software

Regulating Learning
Learn how to set goals and ask for help

Supporting Effort
Identify ways to support students' motivation, engagement, and goal-setting

Supporting Literacy
Access resources to support students' reading strategies, vocabulary, and reading comprehension

Building Mentor Understanding

Open All | Collapse All

Math Content Standards

LONG + LIVE + MATH At Home: Learning Library

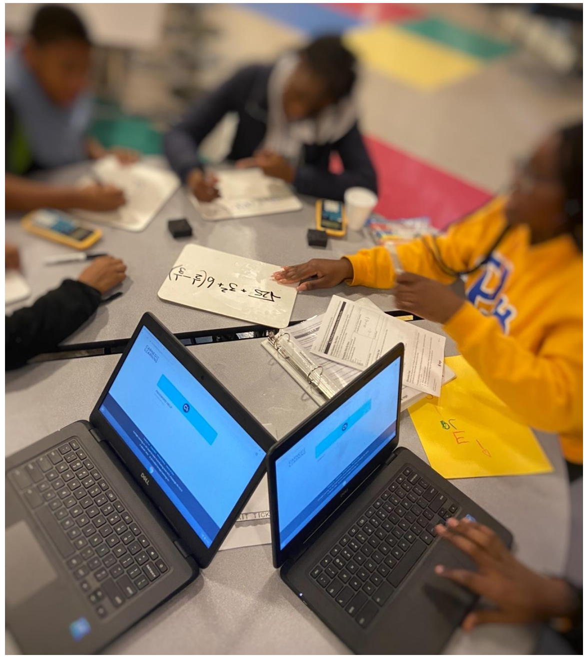
PLUS supports out-of-school tutoring with & without tech

CUE Ready to Learn /PPS	Boys and Girls Club of Western Pennsylvania	Homewood Children's Village

The Ready to Learn Program

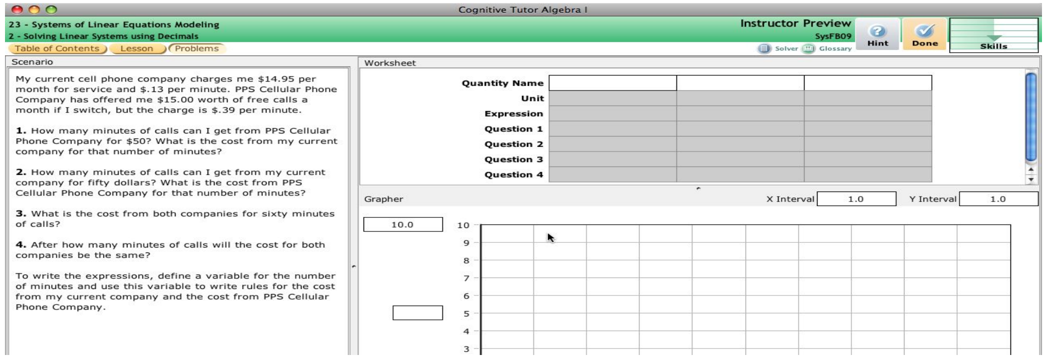
- After-school & summer math mentoring
- Blends tutors & technology to create an engaging learning experience for students
- Tutors are Pitt & CMU undergraduates

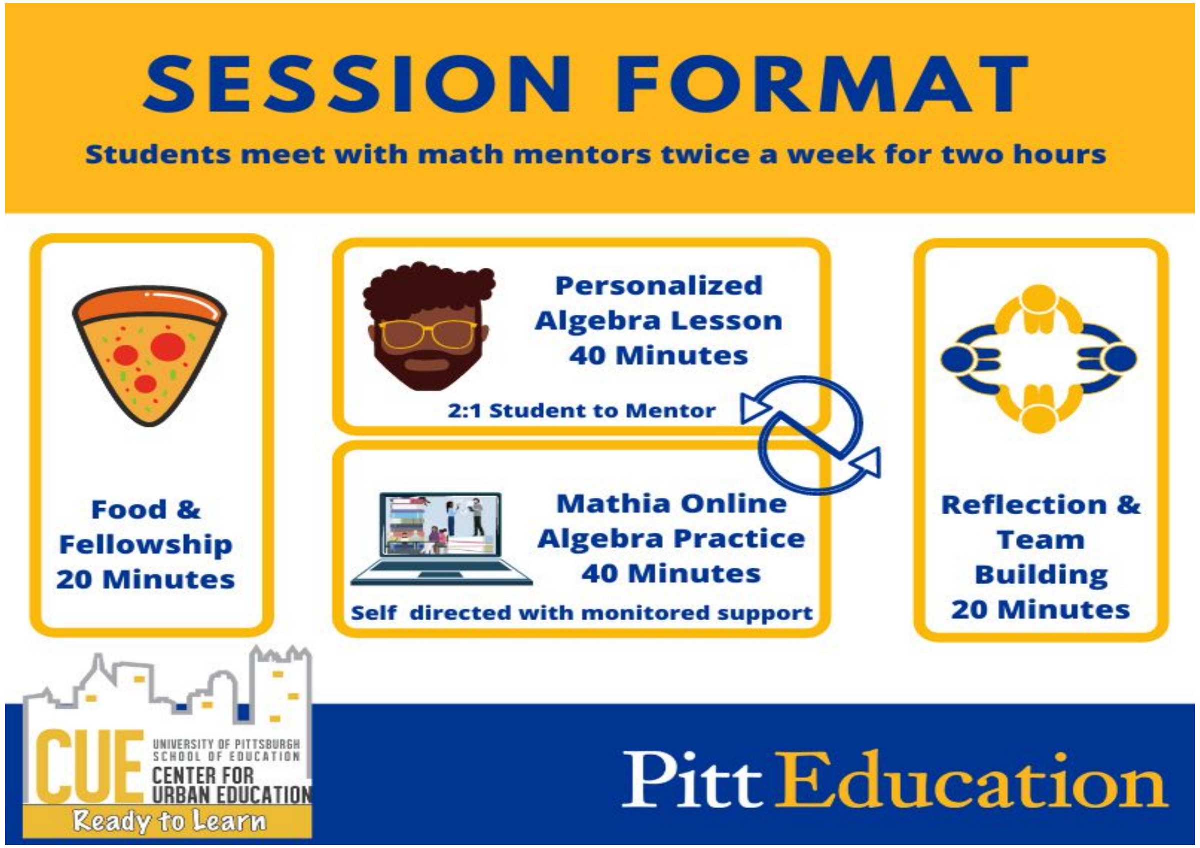
CUE
UNIVERSITY OF PITTSBURGH
SCHOOL OF EDUCATION
CENTER FOR
URBAN EDUCATION
Ready to Learn



Ready to Learn *in Person*

Mathia (Cognitive Tutor)





Quasi-Experimental Evaluation

Compare Math Test Scores (RIT) pre & post

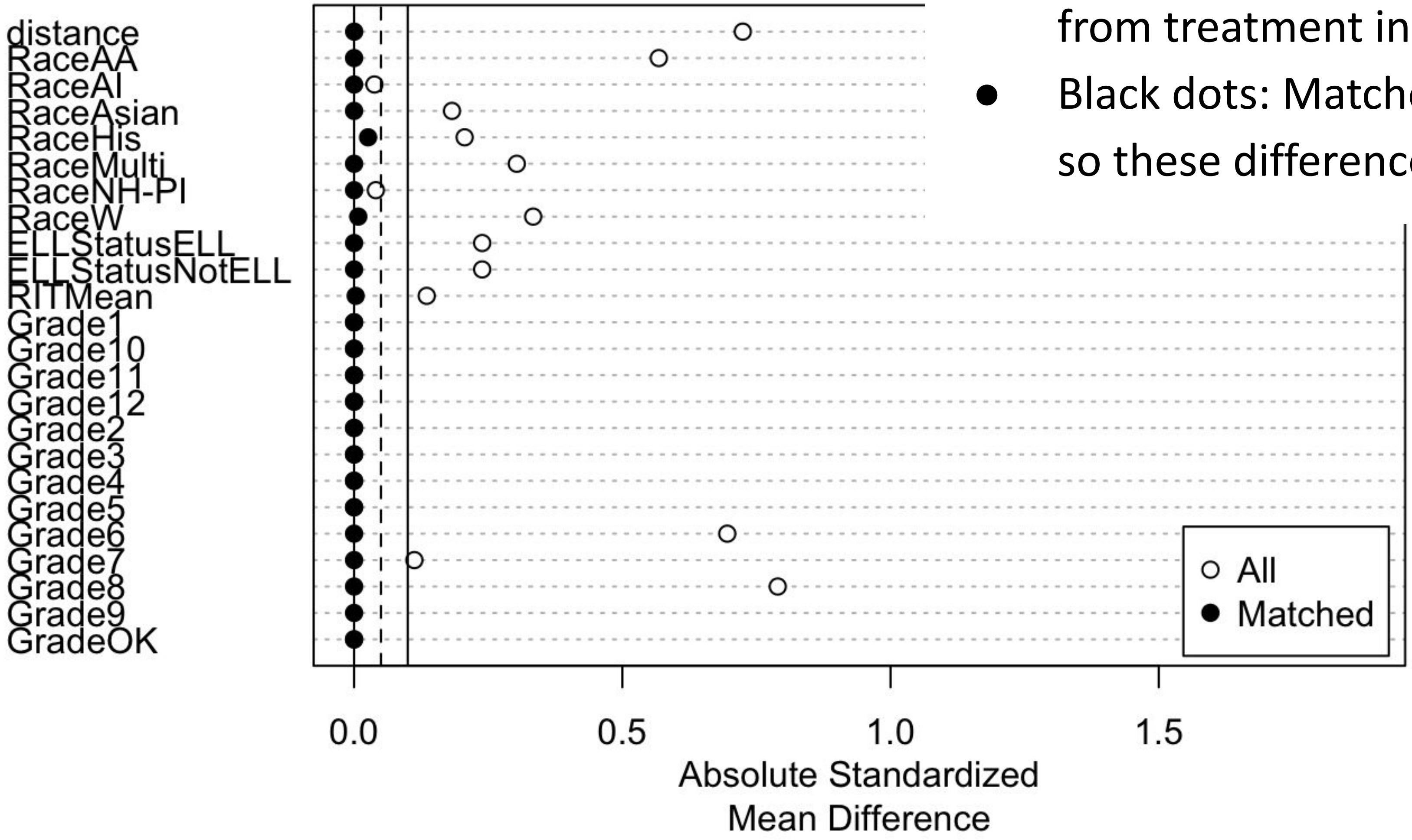
- *Pre*: Fall & Winter (2019-2020) RIT Score
- *Between*: RtL/PL2 for treatment,
  Usual schooling & *pandemic* for all
- *Post*: Fall & Winter (2020-2021) RIT Score

Demographics – 3 local urban schools

- Mostly grades 6-7; 52% female
- 80% black
  8% brown
  12% white

Propensity Matching

- Start: All students at same schools
- White dots: Those students are different from treatment in many ways
- Black dots: Matched students are selected so these differences go to 0



Results: A nonequivalent dependent variables* test shows *no effect* on *reading scores*

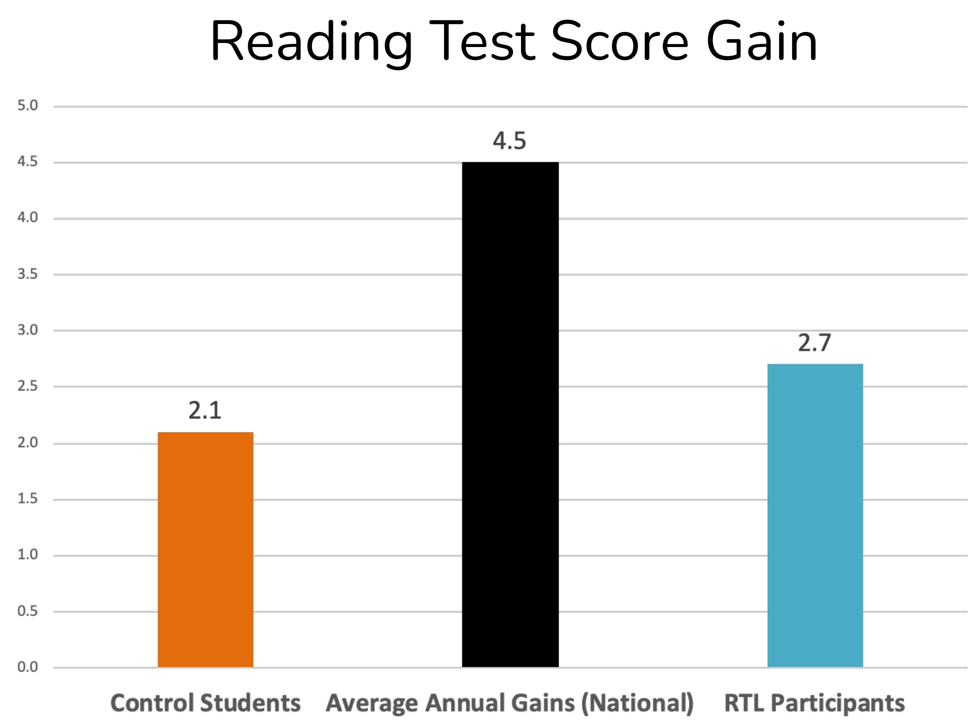
Both groups learn less than "normal"

No difference in amount of learning

RtL students are not simply better learners

Reading Score Difference Results

Variable	***Estimate***	***SE***	***df***	***t***	***p***
(Intercept)	206.55	0.85	492.02	244.08	**0.000*** **
Treatment	1.43	2.13	492.02	0.67	0.051
PostTest	2.05	0.49	417.00	4.20	**0.000*** **
Treatment x PostTest	0.61	1.23	417.00	0.50	0.619



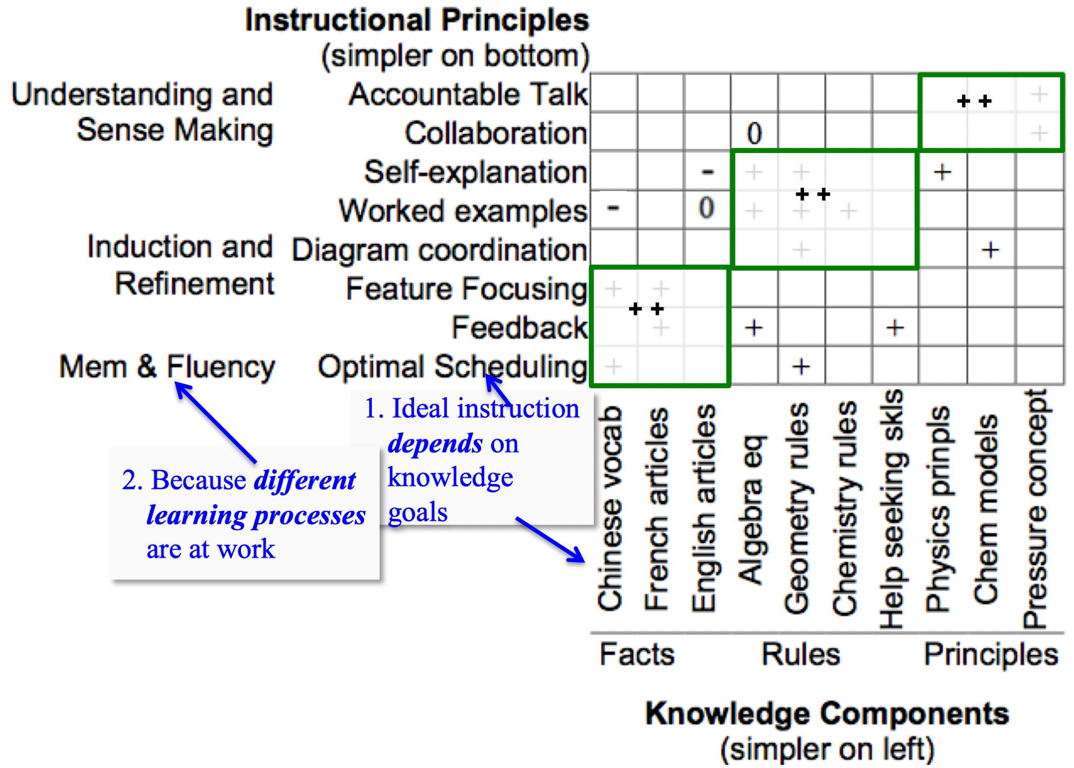
* Trochim & Donnelly (2007). The Research Methods Knowledge Base: 3rd edition.

Knowledge-Learning-Instruction (KLI) Framework

Interactive task-based practice based on discovered **knowledge components** of transfer supports ***memory, induction, & sense-making*** to yield

a) better outcomes than passive alternatives

b) more consistent outcomes for *all students?*



Koedinger, Corbett & Perfetti. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*.

Online tutoring systems provide favorable learning conditions

- Deliberate practice features: Well-tailored tasks, feedback, examples, explanations, repetition, variability
  - Mastery-based repetition
    - Well-tailored tasks ~ Bayesian Knowledge Tracing ~ ZPD
- Supported by theories of skill/schema acquisition
  - Skill acquisition theories: ACT-R, Soar, Apprentice Learner
  - Inductive schema acquisition: Gick & Holyoak …
  - Example source, analogy-based: Genter et al …

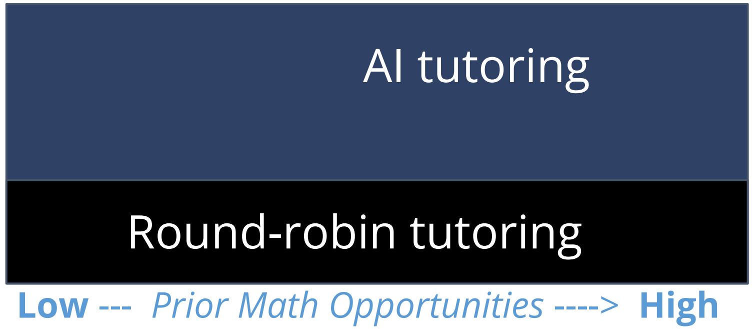
End Extras

Is a 2nd example that connects to an experiment showing benefit of KC model improvement for learning

- Draw from Ran Liu paper

Lean implementation at Site 1

	Site 1
EdTech	**IXL**
Dashboard	**No**
Tutoring strategy	**Round robin**
Remote tech	**Zoom**
%Black or brown	**95%**
%Low income	**96%**



Rapid monitoring of impact of tutoring at Site 1

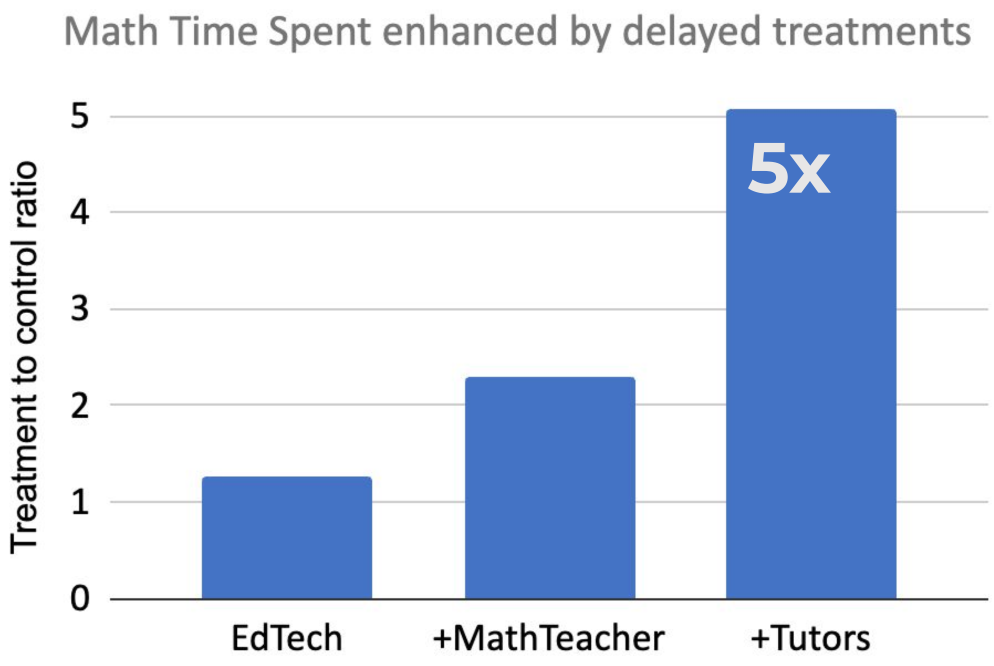
	Fall	**Early Spring**	**Late Spring**
Control	EdTech	EdTech	EdTech
DelayedTreatment	EdTech	EdTech+MathTeacher	EdTech+MathTeacher+Tutoring

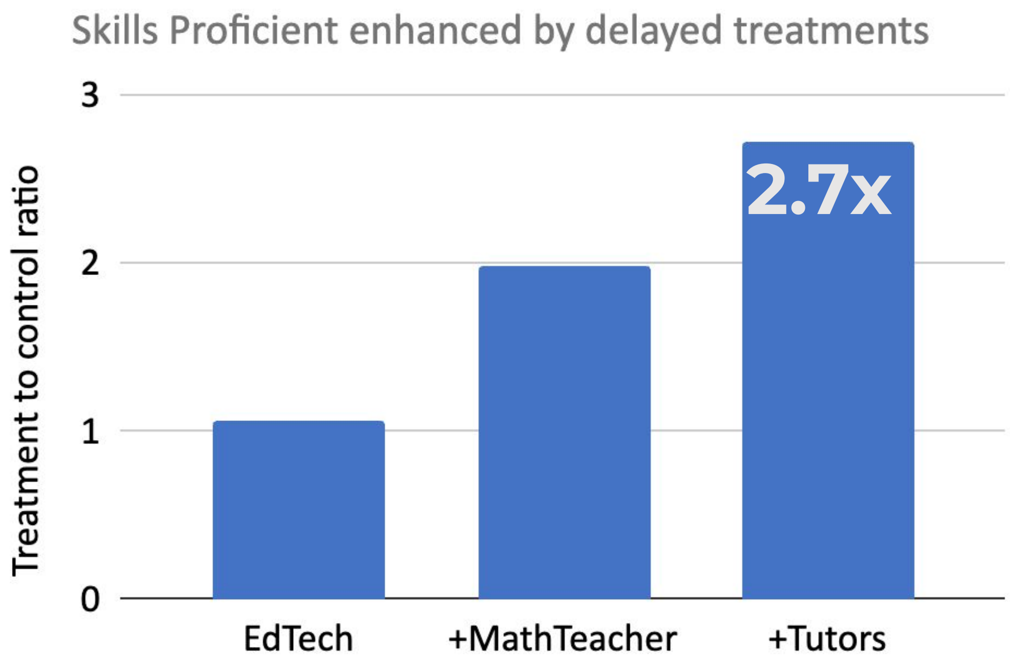
Quasi-experimental methods:

- Interrupted time series
- Non-randomized control

Tutoring increases math effort & progress

	Fall	Early Spring	Late Spring
Control	EdTech	EdTech	EdTech
DelayedTreatment	EdTech	EdTech+MathTeacher	EdTech+MathTeacher+Tutoring





Mixed effects regression for reliable inference

*Time_spent ~ Pretest + Time_period + Student_group + Tutoring * MathTeacher + Pretest:Tutoring + Pretest:MathTeacher + (1 | studentID) + (1 | week)*

Predictors	Estimate	p-value
Pretest	0.015	0.755105
Early_spring (vs. Fall)	-0.121	0.282359
Late_spring (vs. Fall)	-0.495	0.000660 ***
Student_group_Control	-0.048	0.613820
MathTeacher (vs. EdTech)	0.487	< 2e-16 ***
Pretest:MathTeacher	0.271	1.05e-09 ***
Tutoring (vs. MathTeacher)	0.202	0.006279 **
Pretest:Tutoring	-0.212	0.000351 ***

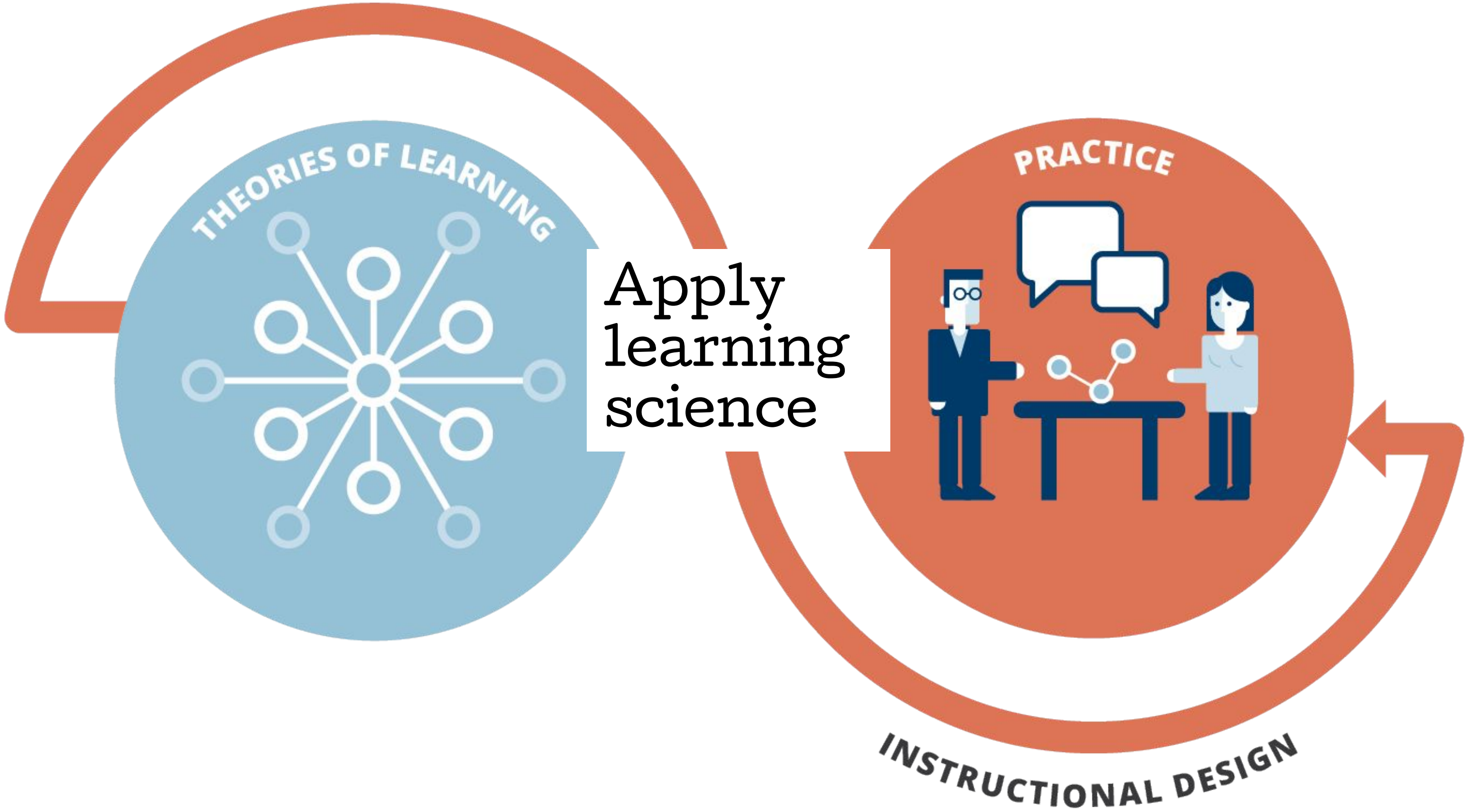
Math Teacher increases EdTech use

Esp. for ***more*** *prepared students*

Tutoring further increases

Esp for ***less*** *prepared students*

Is this what we should be doing?



Simply Applying Learning Science
Is Not Enough!

- Only ~10% of US Dept of Ed funded randomized controlled trials (RCT)
  find positive effects
  - 11/90 IES RCTs succeeded
    - See 2013 review at Coalition4evidence.org
- Recommendations differ
  - especially from different disciplines:
    Cog Neuro, Psychology, Ed Psych, Education

Can’t just *apply*
learning science,

must *do*
learning science

Overview

Better predictions are ***not yet trustworthy*** because they may not have causal power to improve learning

Trustworthy learning analytics produce insights that yield enhanced student learning

- OK analytics: Better prediction. ~70%?
- Trustworthy: Better prediction => Insight & redesign => Close-the-loop experiment

=11% <<11%??

Loop 1

- Analysis: What student choices predict learning outcomes?
- Close-the-loop: Active doing vs. passive reading learning experiment

Loop 2

- Analysis: What student differences account for learning outcome differences?
- Close-the-loop: Hybrid human-AI tutoring experiment

MOOC Psychology Course features

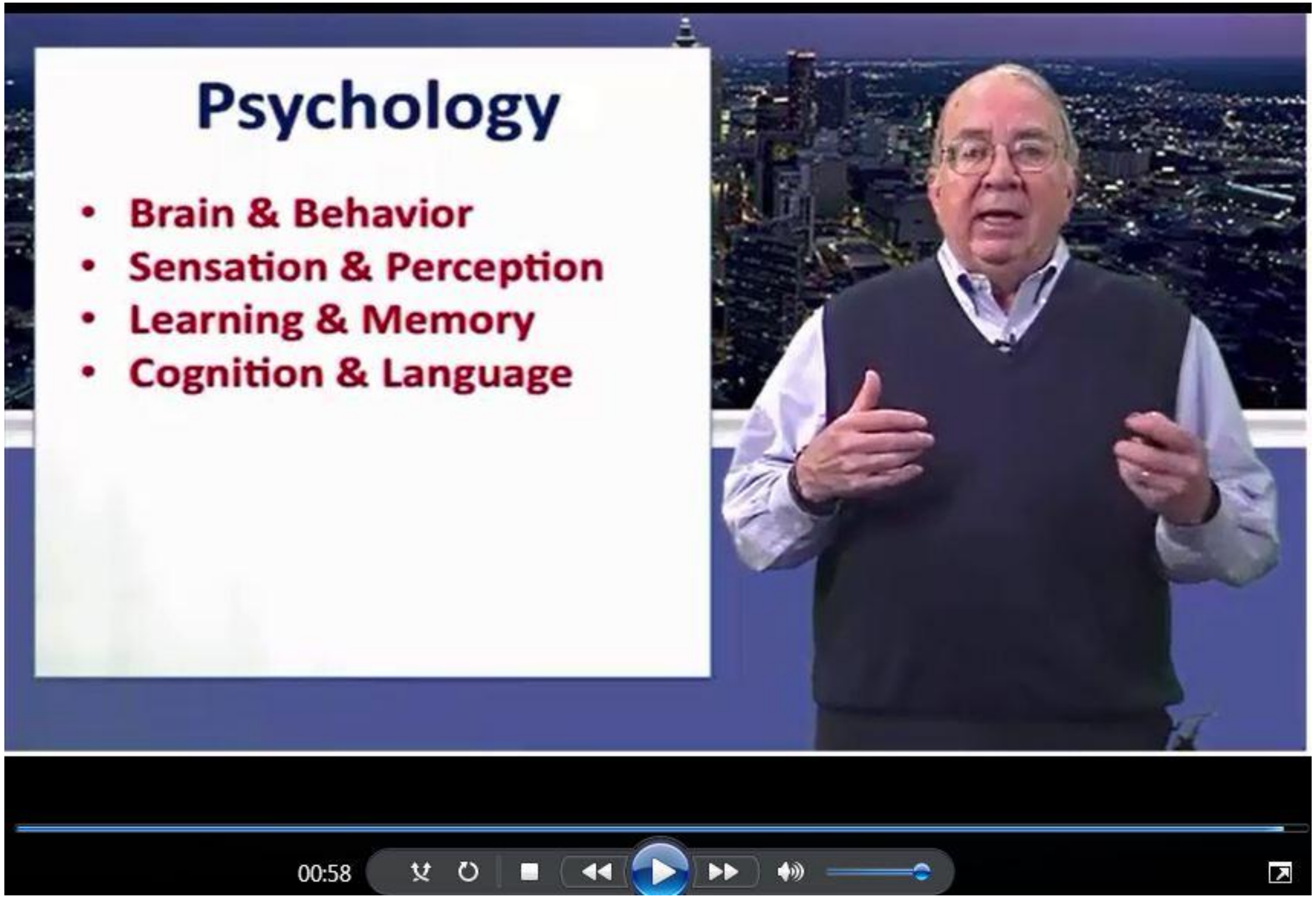
Coursera: Passive/declarative	OLI: Active/interactive
Video lectures Discussion forums Writing & surveys Unit quizzes Final	Readings Interactive practice - feedback targets misconceptions - hints support students

Carnegie Mellon University

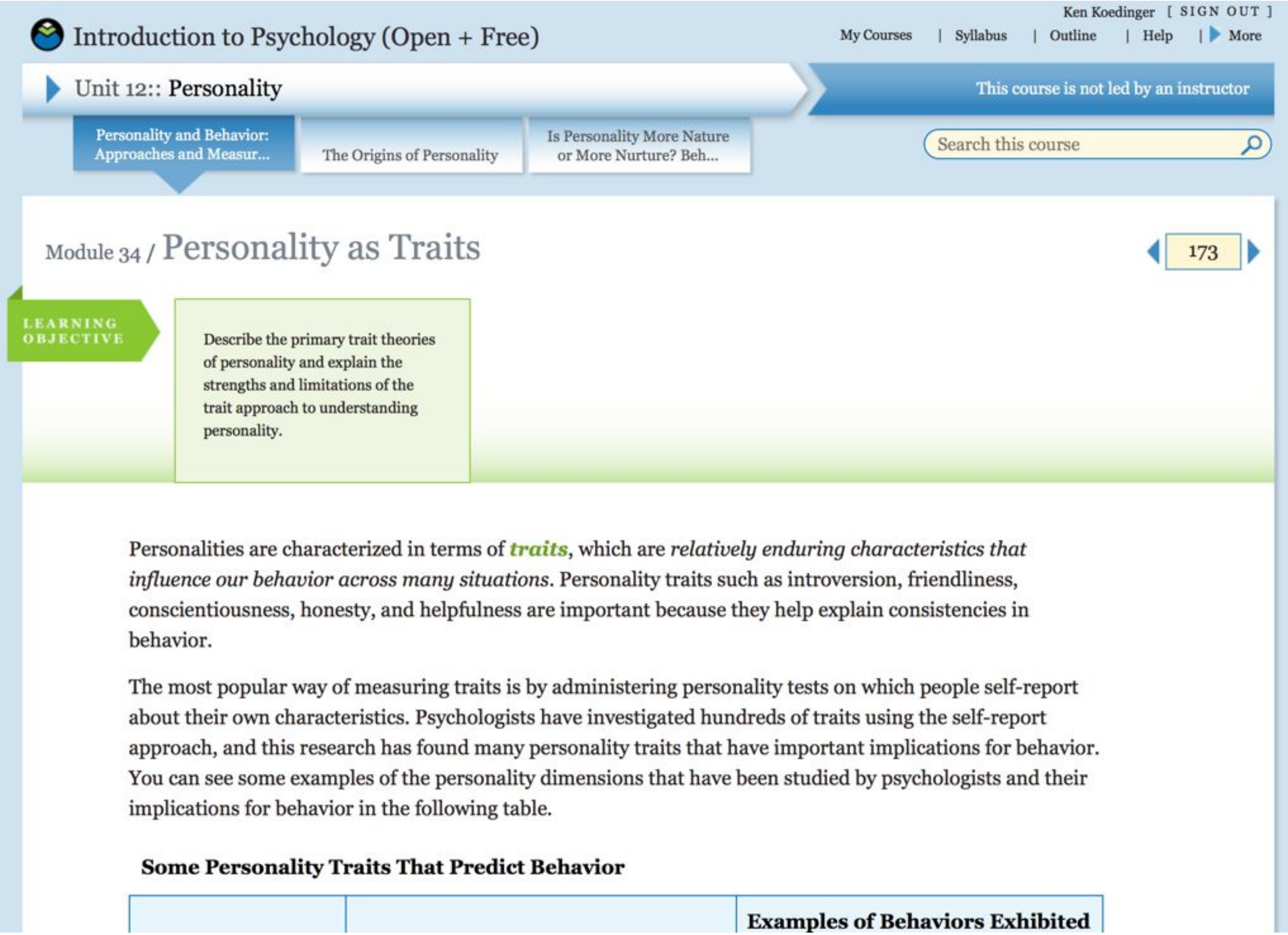
Open Learning Initiative

Transforming higher education through the science of learning.

MOOC video example – “watching”



OLI content page – "reading"



Ken Koedinger [SIGN OUT]

Introduction to Psychology (Open + Free)

My Courses | Syllabus | Outline | Help | More

Unit 12:: Personality

This course is not led by an instructor

Personality and Behavior: Approaches and Measur... | The Origins of Personality | Is Personality More Nature or More Nurture? Beh...

Search this course

Module 34 / Personality as Traits

173

LEARNING OBJECTIVE

Describe the primary trait theories of personality and explain the strengths and limitations of the trait approach to understanding personality.

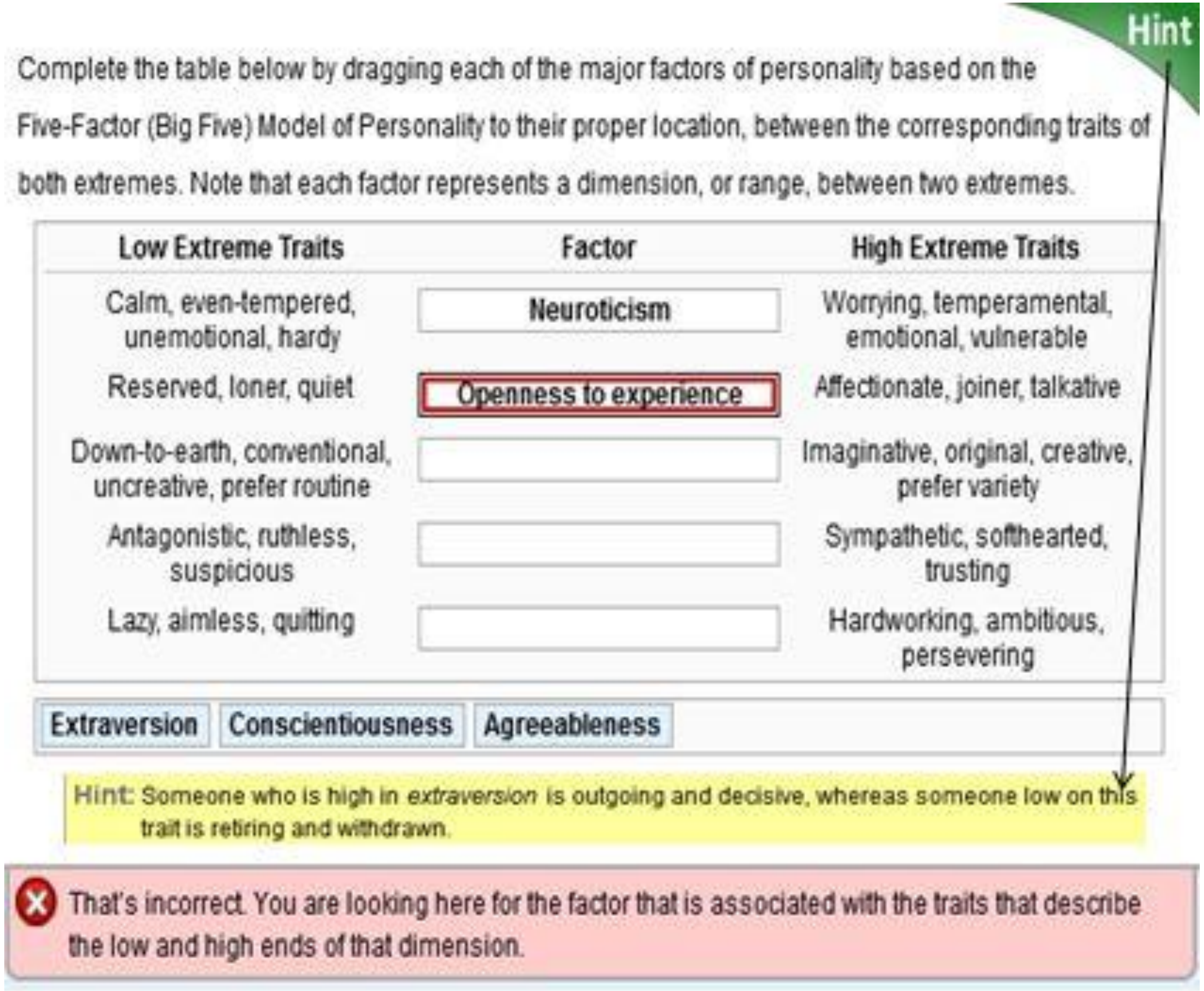
Personalities are characterized in terms of ***traits***, which are *relatively enduring characteristics that influence our behavior across many situations*. Personality traits such as introversion, friendliness, conscientiousness, honesty, and helpfulness are important because they help explain consistencies in behavior.

The most popular way of measuring traits is by administering personality tests on which people self-report about their own characteristics. Psychologists have investigated hundreds of traits using the self-report approach, and this research has found many personality traits that have important implications for behavior. You can see some examples of the personality dimensions that have been studied by psychologists and their implications for behavior in the following table.

Some Personality Traits That Predict Behavior

		Examples of Behaviors Exhibited

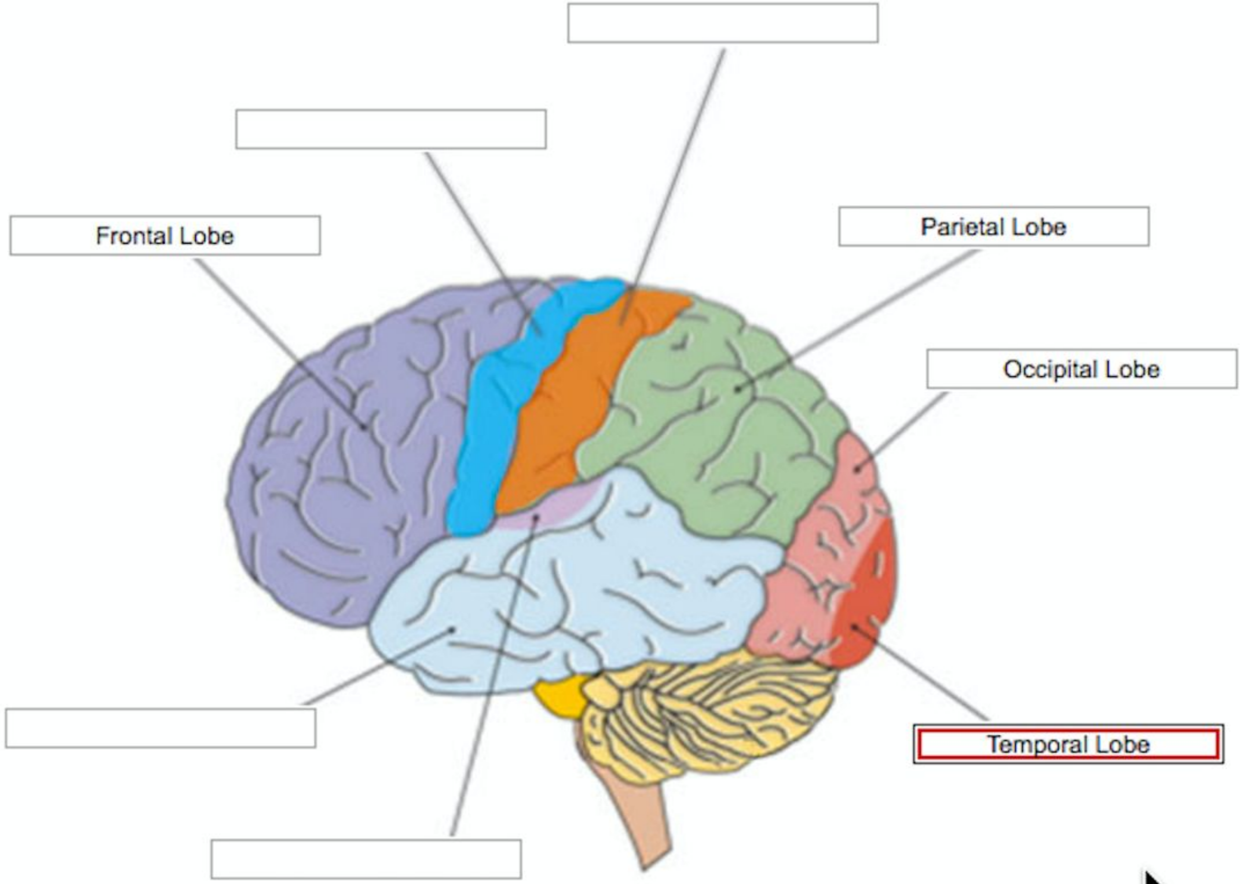
OLI activity – “doing” = *practice with feedback*



(a) OLI Activity

Hint

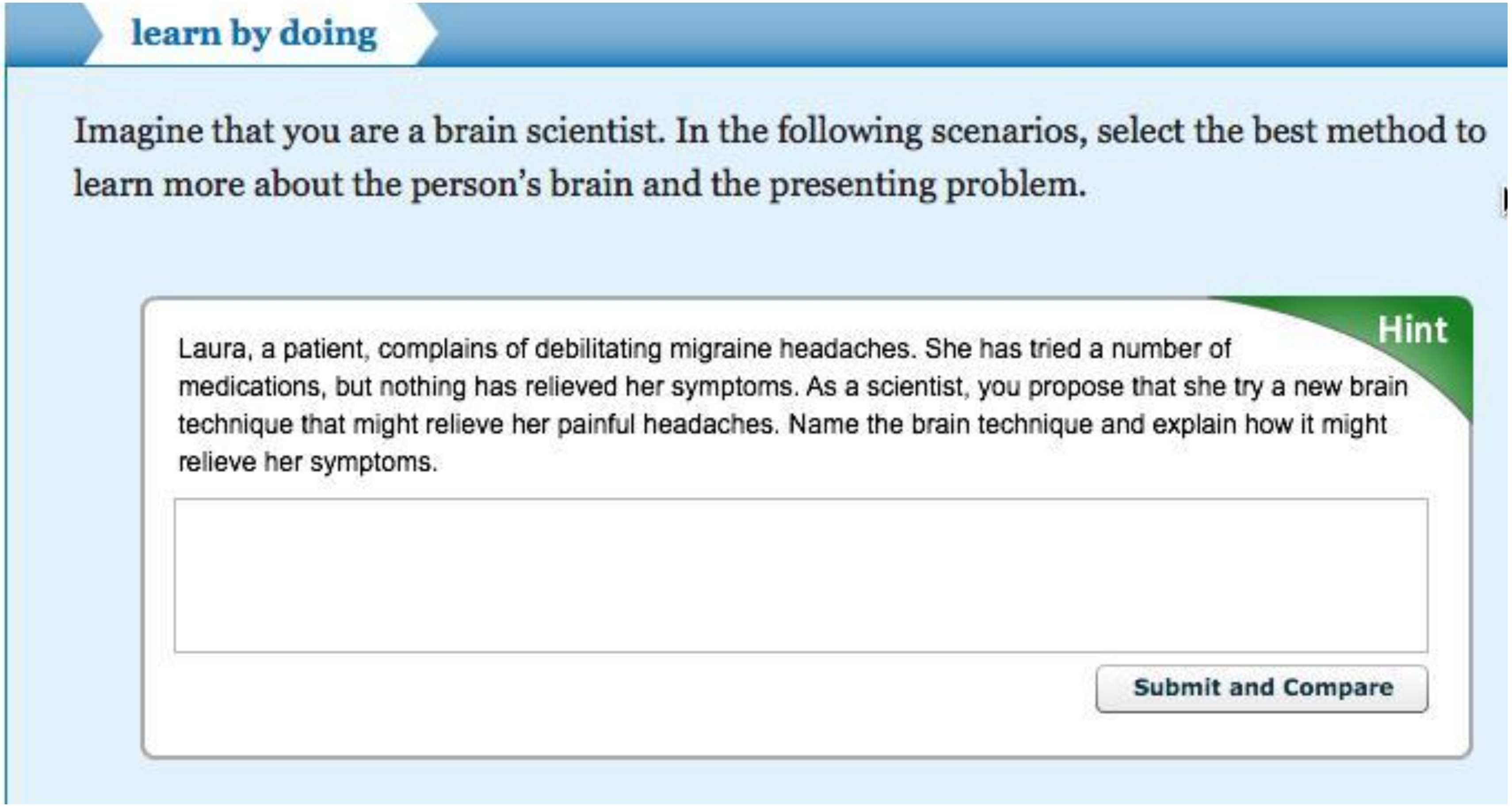
Complete the following diagram by dragging the labels at the bottom into the appropriate spots on the diagram.



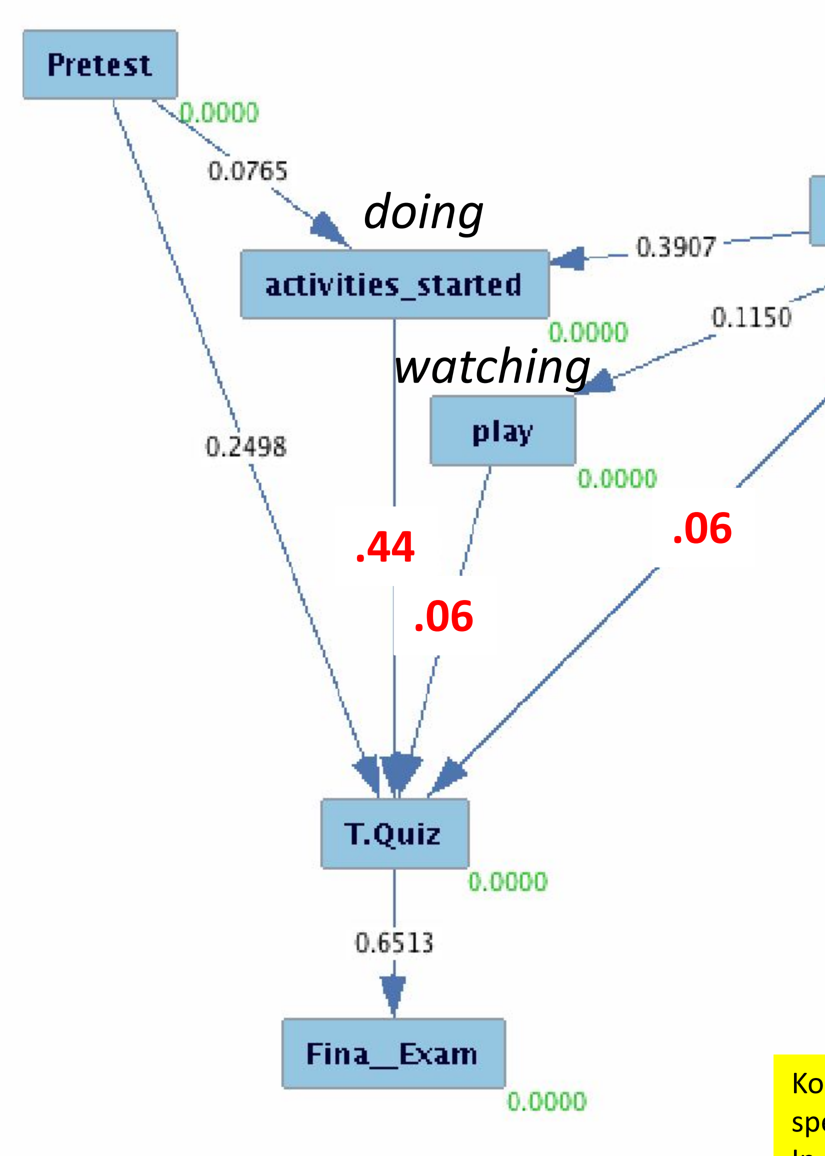
Motor Cortex | Somatosensory Cortex | Visual Cortex | Auditory Cortex

That's incorrect. The temporal lobe is one of the four major sections of each hemisphere of the cerebral cortex.

Open-ended “submit and compare”



Causal inference analysis (using “Tetrad”)



Regression weights in standard deviation units

1 sd increase of *doing* leads to .44 increase in Quiz

Learning by doing > 6x better than learning by watching!!

Koedinger, Kim, Jia, McLaughlin, & Bier (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the Second ACM Conference on Learning at Scale*.

Doer effect found in other courses (N = 12K students)

	Doing to reading effect ratio	
	Quizzes	**Final Grade**
Info Systems	5.2	2.2
Biology	5.0	3.1
Statistics	∞	16.4
Psych	8.5	7.7
Psy MOOC	6.8	4.8

Doing to reading effect ratio: 2.2x to >16x

– Median is 6x

Koedinger, McLaughlin , Jia, & Bier (2016).
Is the Doer Effect a Causal Relationship?
In *Proceedings of LAK*.

Doer effect found by other researchers

The Doer Effect at Scale: Investigating Correlation and Causation Across Seven Courses

Rachel Van Campenhout, Bill Jerome, Jeff Dittel & Benny Johnson

LAK 2023

Do students who **do more** in unit 4 than unit 7 **learn more** in unit 4 than in unit 7?

Koedinger, McLaughlin , Jia, & Bier (2016). Is the Doer Effect a Causal Relationship? In *Proceedings of LAK*.

Analytic method: Mixed effects regression

Learning method	Location	Normalized Estimate	Pr(>|t|)
	Before	0.143	< 0.00 ***
Doing	*Within*	0.181	< 0.00 ***
	After	0.078	0.00 ***
	Before	0.008	0.56278
Reading	*Within*	0.010	0.22116
	After	-0.013	0.30410
	Before	0.054	0.00012 ***
Watching	*Within*	0.025	0.00180 **
	After	0.033	0.01343 *

Yes!
Increasing evidence that *active practice **causes*** more learning than *passive reading or watching*

But:

- Third variable explanations are still possible
- Positive association with more doing & watching *after* a unit

Close-the-loop experiment needed!

Summary

Better predictions are ***not yet trustworthy*** because they may not have causal power to improve learning

Trustworthy learning analytics produce insights that yield enhanced student learning

- OK analytics: Better prediction. **~70%?**
- Trustworthy: Better prediction => Insight & redesign => Close-the-loop experiment

=11% **<<11%??**

Loop 1

- Analysis: *More doing is 6x more associated with learning than more watching or reading*
- Close-the-loop: Active doing vs passive reading learning experiment

Loop 2

- Analysis: Learning rate regularity
- Close-the-loop: Hybrid human-AI tutoring experiment

Overview

Better predictions are ***not yet trustworthy*** because they may not have causal power to improve learning

Trustworthy learning analytics produce insights that yield enhanced student learning

- OK analytics: Better prediction. ~70%?
- Trustworthy: Better prediction => Insight & redesign => Close-the-loop experiment

=11% <<11%??

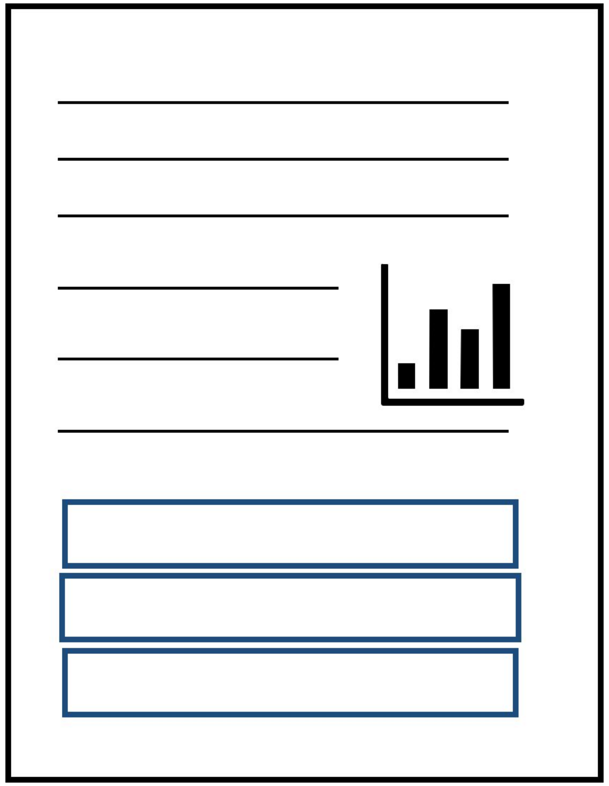
Loop 1

- Analysis: *More doing is 6x more associated with learning than more watching or reading*
- Close-the-loop: Active doing vs passive reading learning experiment

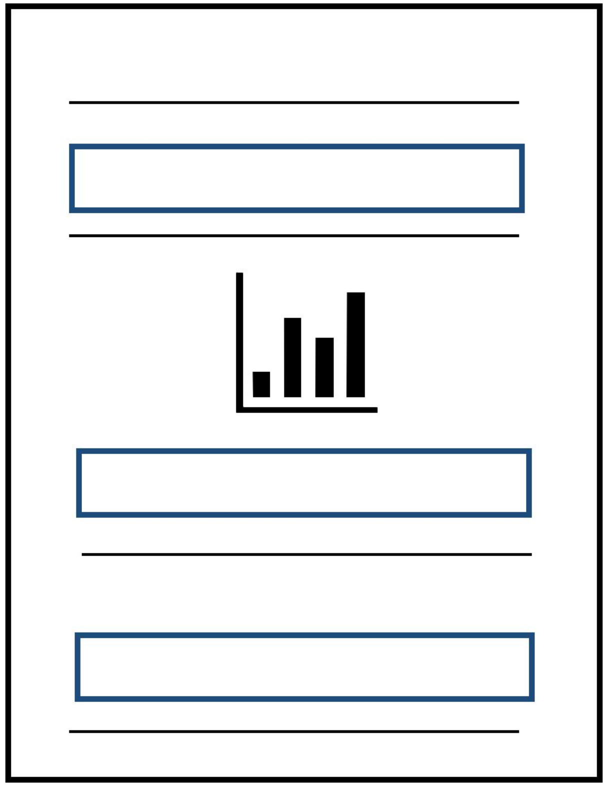
Loop 2

- Analysis: What student differences account for learning outcome differences?
- Close-the-loop: Hybrid human-AI tutoring experiment

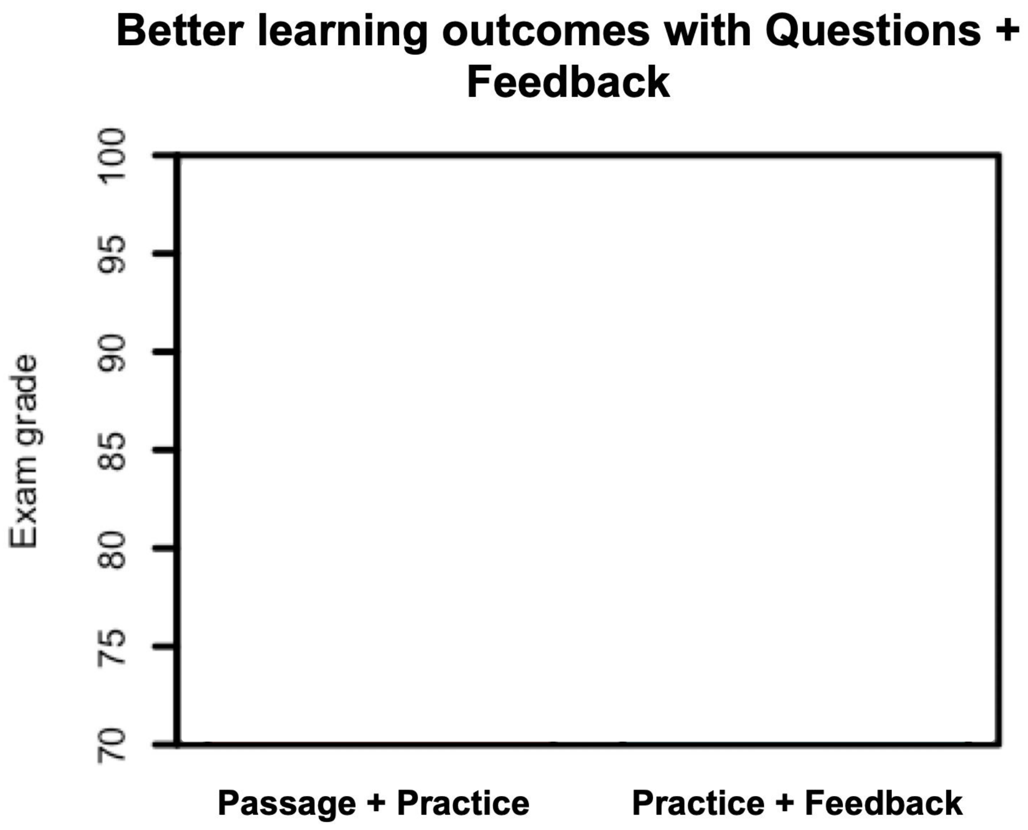
Can activities without prior reading improve learning in a classroom?



Passage+Practice



Practice+Feedback



N = 75 students enrolled in Intro Psych (within subject design)

Carvalho, Manke, & Koedinger (2018; CogSci proceedings)

Summary

Better predictions are ***not yet trustworthy*** because they may not have causal power to improve learning

Trustworthy learning analytics produce insights that yield enhanced student learning

- OK analytics: Better prediction. ~70%?
- Trustworthy: Better prediction => Insight & redesign => Close-the-loop experiment

=11% <<11%??

Loop 1

- Analysis: *More doing is 6x more associated with learning than more watching or reading*
- Close-the-loop: *Replace up-front reading with up-front doing*

Loop 2

- Analysis: What student differences account for learning outcome differences?
- Close-the-loop: Hybrid human-AI tutoring experiment