Crowdsourcing Cognitive Models for Assessment, Tutoring, and In-Game Support

Ken Koedinger

Human-Computer Interaction & Psychology

Carnegie Mellon University

Talk for
Crowdsourcing Personalized Online Education
July 2012

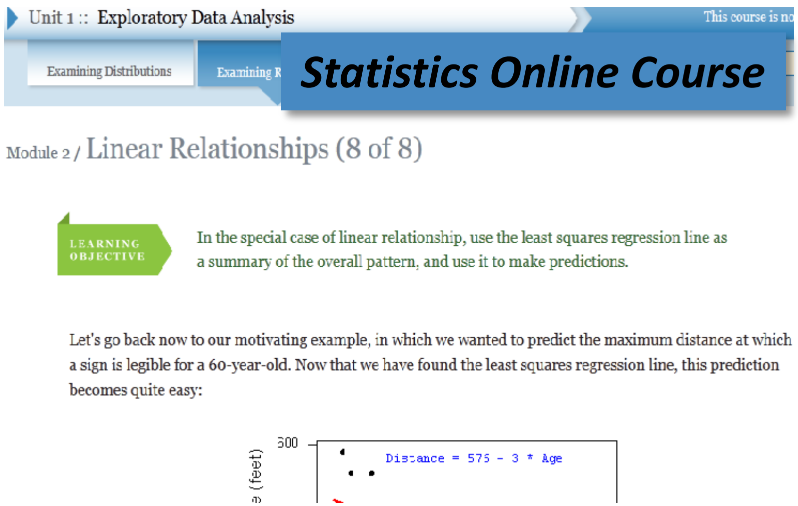
Argument Overview

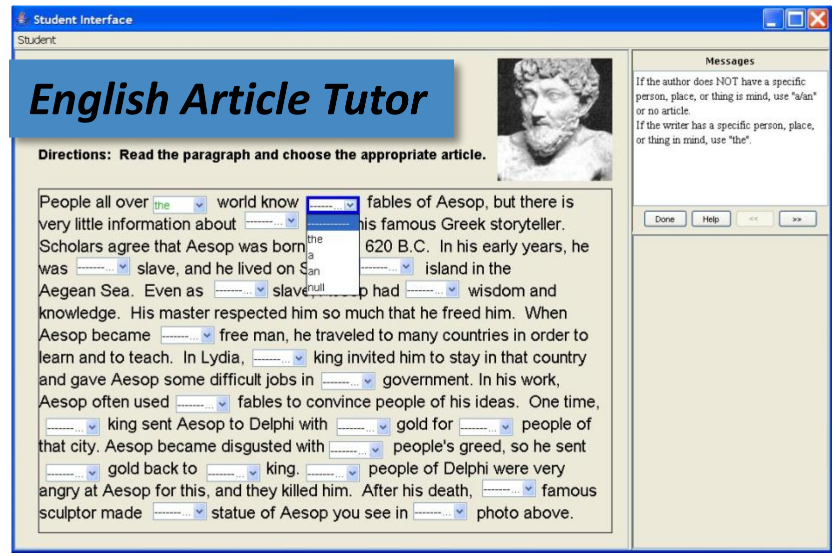
- Student models drive effective instruction
- Data can be used to create better models
  - LearnLab's DataShop: Datasets, learning curve visualizations, model comparison leaderboards, data mining algorithms
- Result: Combined human & machine intelligence discovers new models across domains
- Practical consequence: Focused redesign to improve instruction

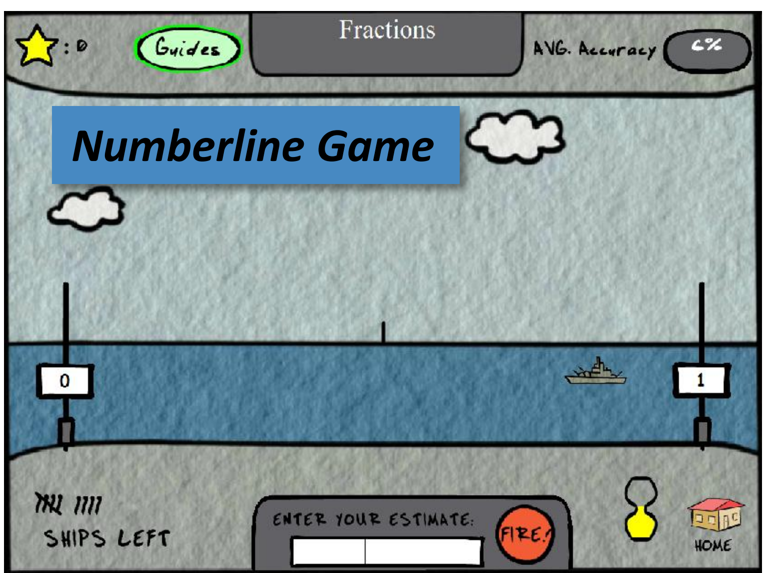
Devil's in the details

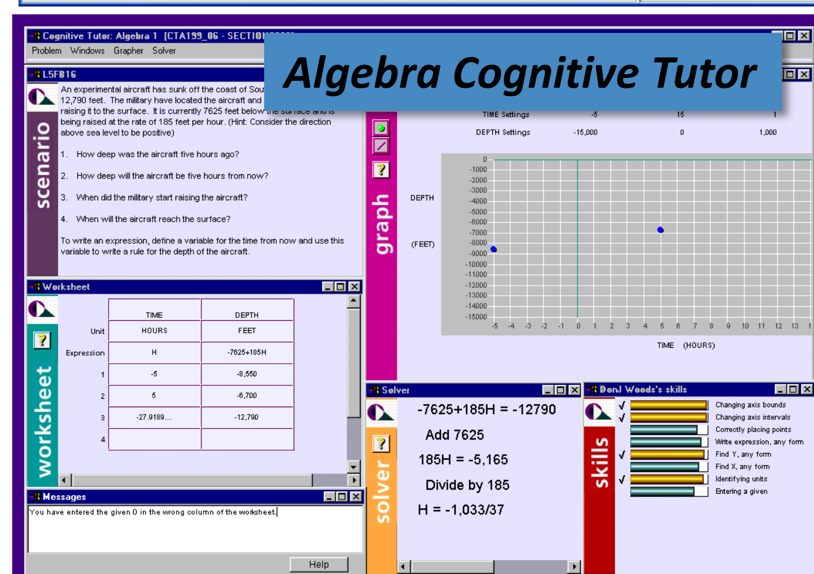
Koedinger, McLaughlin, & Stamper (2012). Automated student model improvement. In *Proceedings of the Fifth International Conference on Educational Data Mining*. [Conference best paper.]

Data from a variety of educational technologies & domains





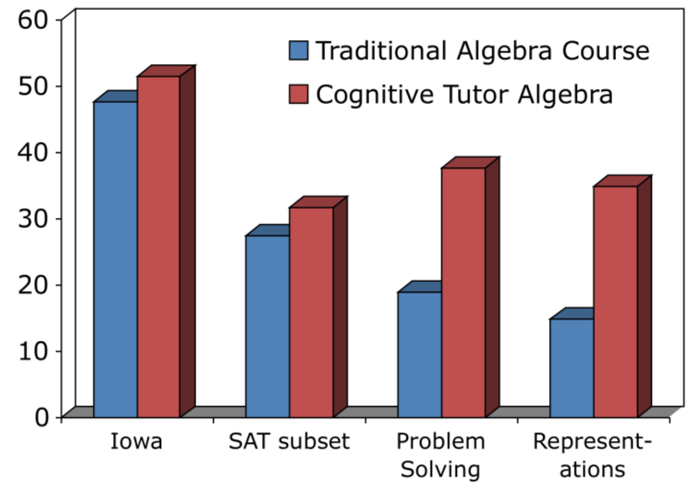




Intelligent tutoring goes to school

Algebra Cognitive Tutor

- Based on computational models of student thinking & learning
- Enhances student learning
- Widespread intensive use
  ~600K students use
  ~80 minutes per week



- 1998 Spin-off:

Carnegie Learning™
THE COGNITIVE TUTOR® COMPANY

Koedinger, Anderson, Hadley, & Mark (1997). Intelligent tutoring goes to school in the big city.

Cognitive Tutor Technology

Use cognitive model to individualize instruction

- Cognitive Model: A system that can solve problems in the various ways students can

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

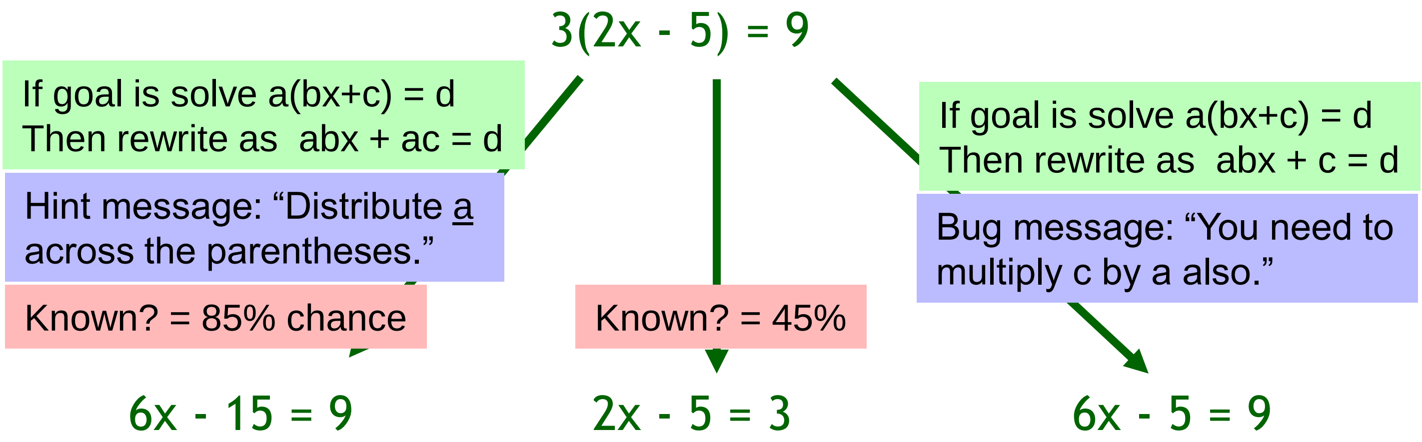
6x - 15 = 9 2x - 5 = 3 6x - 5 = 9

- Model Tracing: Follows student through their individual approach to a problem -> context-sensitive instruction

Cognitive Tutor Technology

Use cognitive model to individualize instruction

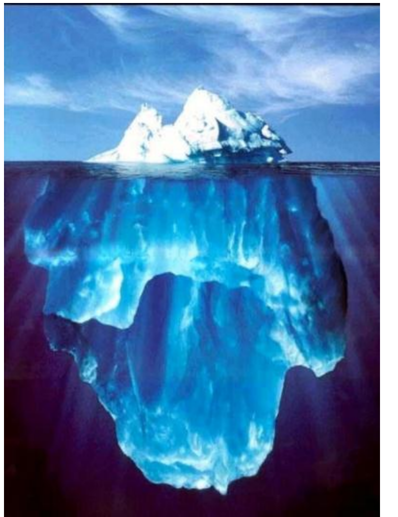
- Cognitive Model: A system that can solve problems in the various ways students can



- Model Tracing: Follows student through their individual approach to a problem -> context-sensitive instruction
- Knowledge Tracing: Assesses student's knowledge growth -> individualized activity selection and pacing

Why data is important to improving student learning

- If we knew everything about students' learning challenges, we would not need data
- But, there is a lot we *do not know* about student learning
- In fact, there's a lot we don't know about our *own* learning
  - You've had lots of experience with the English language
  - You might say you know English
  - But, *do you know what you know?*



Which kind of problem is most difficult 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?

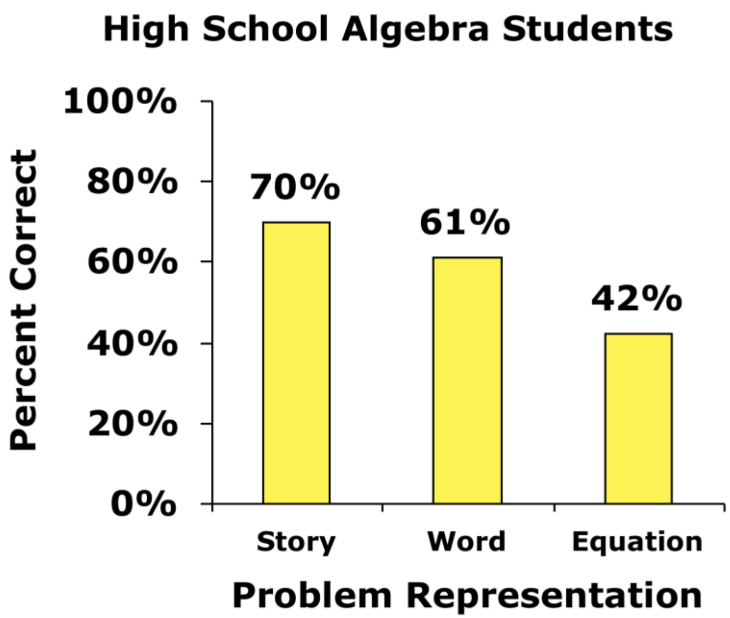
Word Problem

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

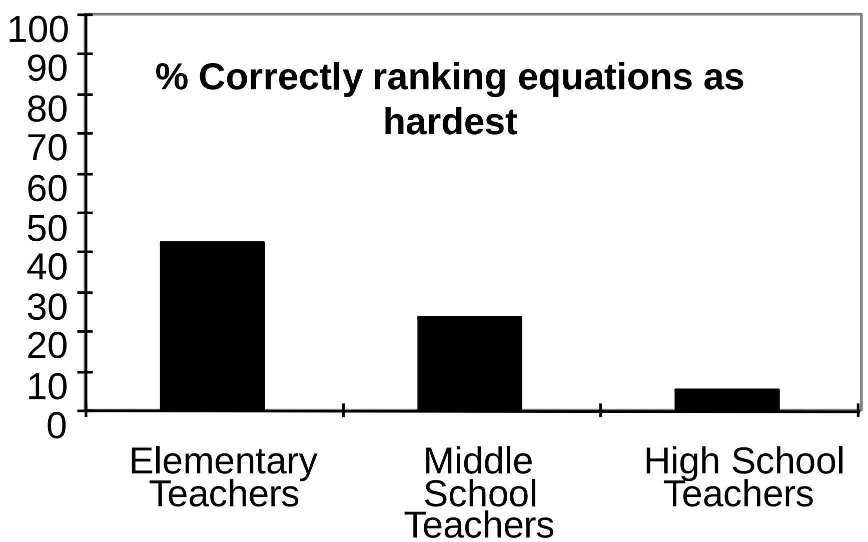
Equation

$x * 6 + 66 = 81.90$

Data contradicts common beliefs of researchers and teachers



Expert Blind Spot!



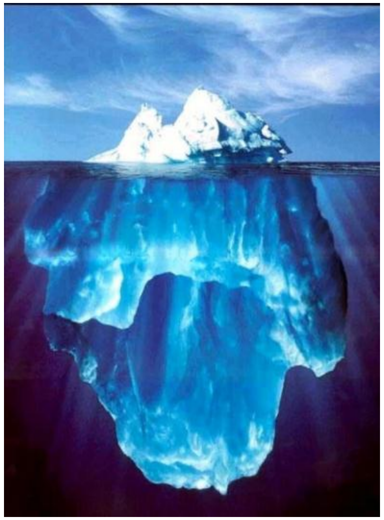
Koedinger & Nathan (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*.

Nathan & Koedinger (2000). An investigation of teachers' beliefs of students' algebra development. *Cognition and Instruction.*

Why have cognitive/student models?

- Drive personalization decisions
  - Better student or cognitive model
    => better feedback & context-sensitive instruction
    => better adaptive problem selection
- Relevant to improving instruction more generally

So, how can we develop & improve student models?



Cognitive Task Analysis

Methods: Expert interviews, think alouds

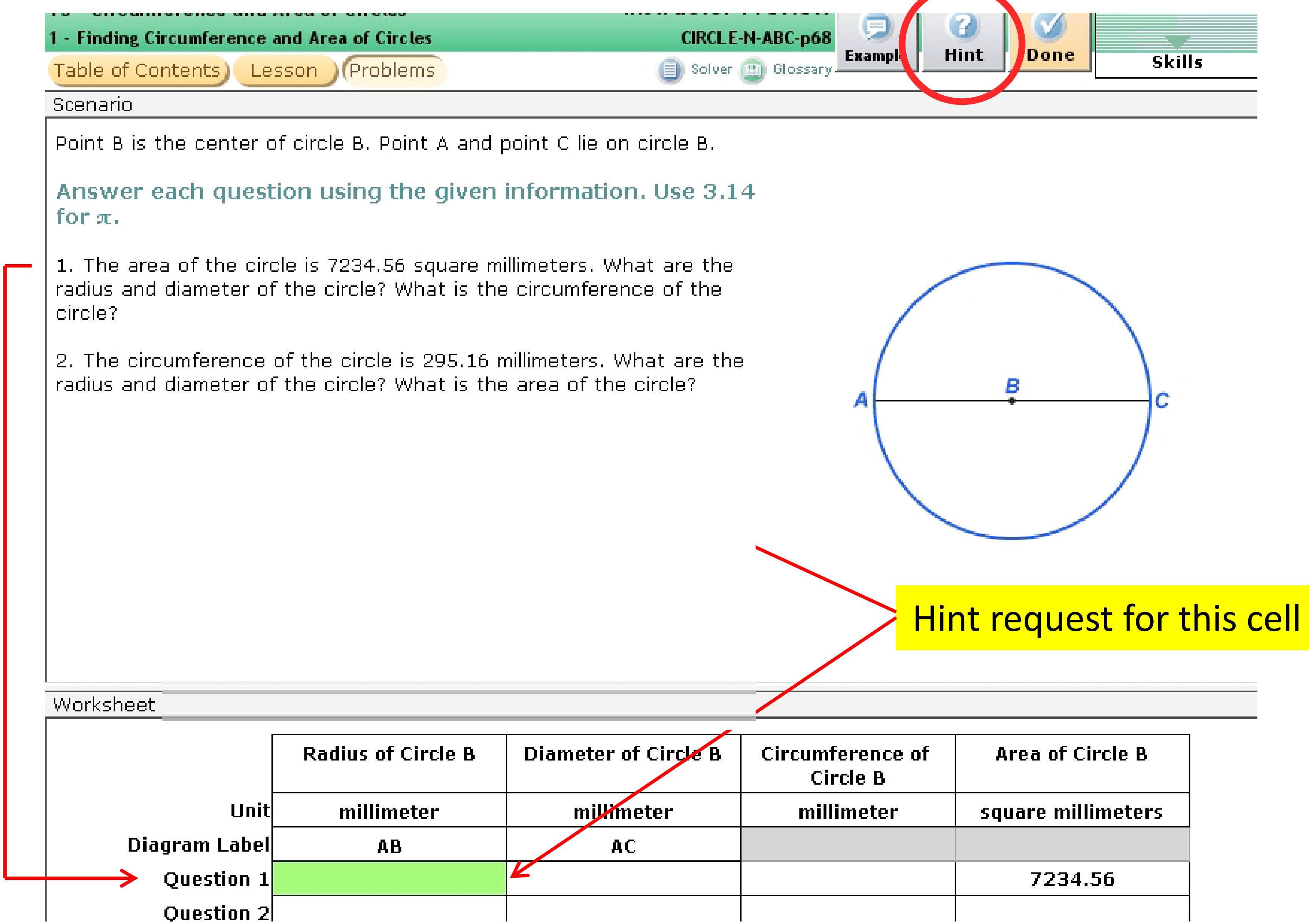
Benefits: Better cognitive models & better instruction, *1.7 sd*! (Clark et al, 2008)

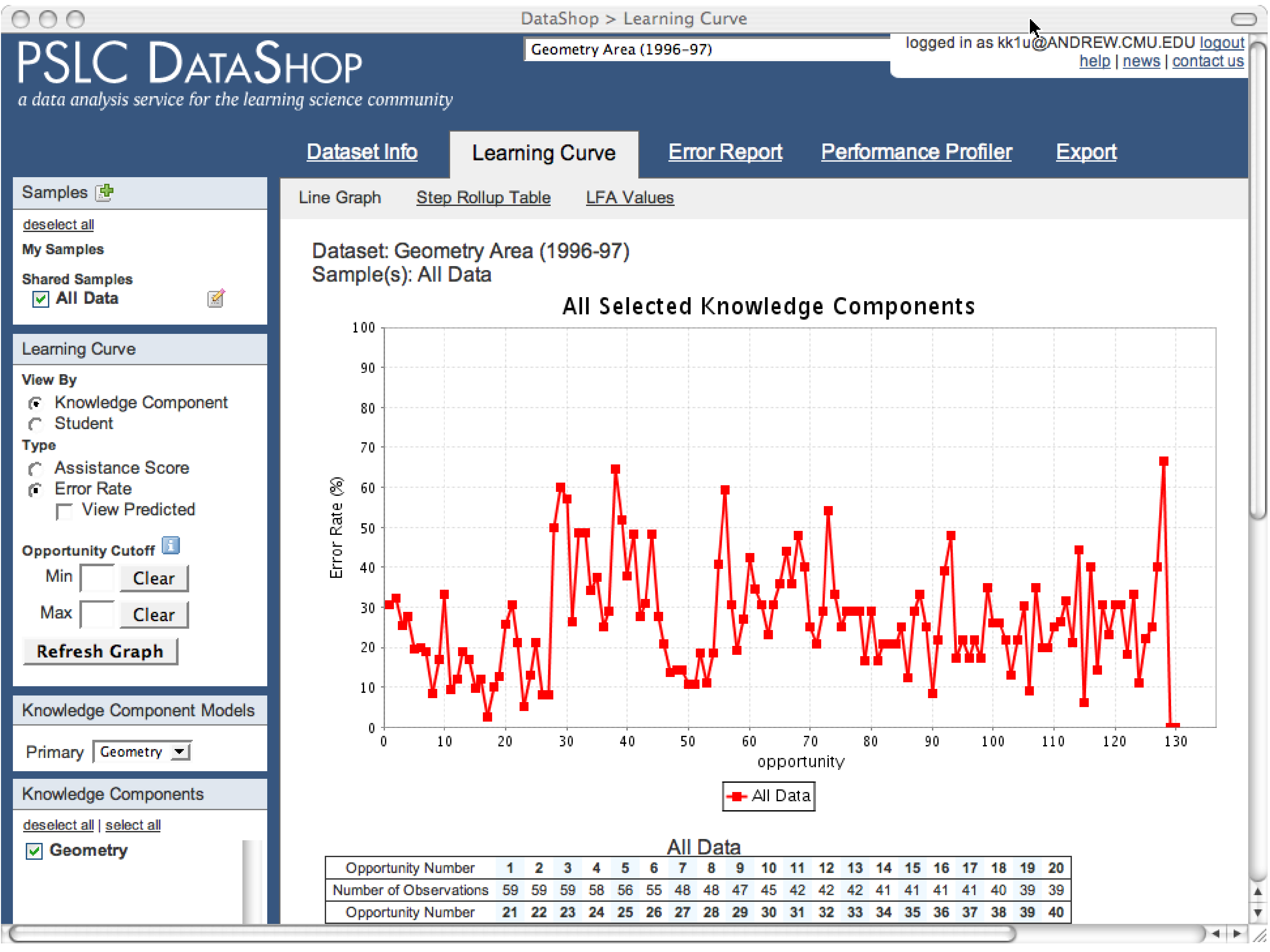
But: Time & expertise intensive

Instead: Use ed tech datasets: learnlab.org/datashop

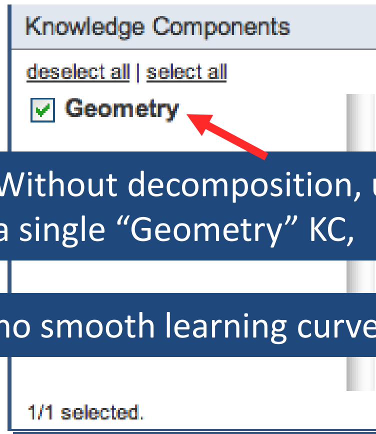
Insight: Build statistical models of student performance: *latent variables* represent *components of symbolic cognitive model*

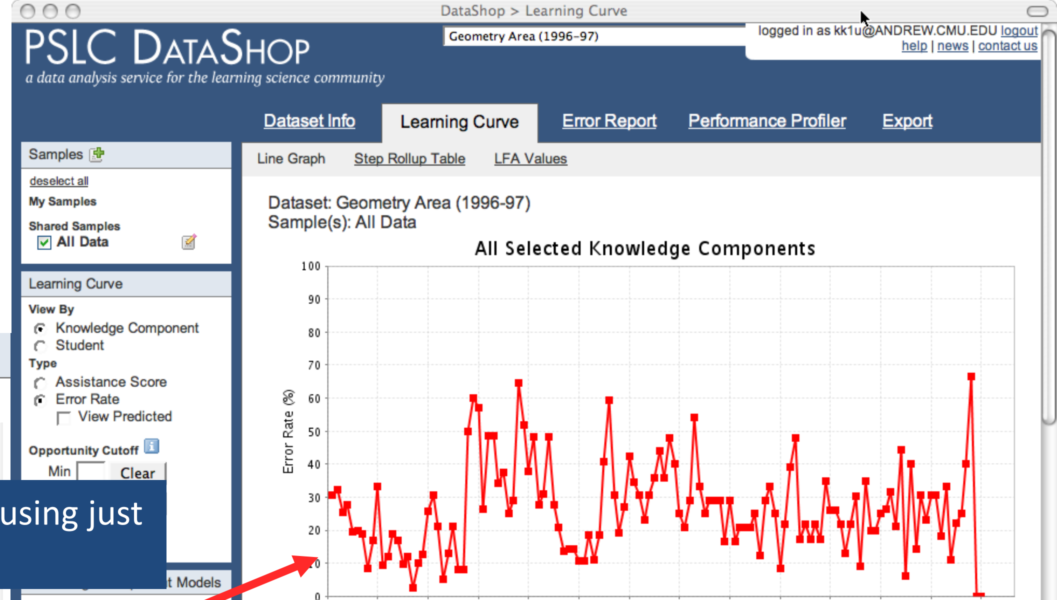
Problem Circle N in the area unit of the Geometry Cognitive Tutor




DataShop > Learning Curve
PSLC DataShop
a data analysis service for the learning science community
Geometry Area (1996-97)
logged in as kk1u@ANDREW.CMU.EDU logout
help | news | contact us
Dataset Info
Learning Curve
Error Report
Performance Profiler
Export
Samples
deselect all
My Samples
Shared Samples
All Data
Learning Curve
View By
Knowledge Component
Student
Type
Assistance Score
Error Rate
View Predicted
Opportunity Cutoff
Min
Clear
Max
Clear
Refresh Graph
Knowledge Component Models
Primary
Geometry
Knowledge Components
deselect all | select all
Geometry
Line Graph
Step Rollup Table
LFA Values
Dataset: Geometry Area (1996-97)
Sample(s): All Data
All Selected Knowledge Components
Error Rate (%)
opportunity
All Data
All Data
Opportunity Number 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Number of Observations 59 59 59 58 56 55 48 48 47 45 42 42 42 41 41 41 41 40 39 39
Opportunity Number 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

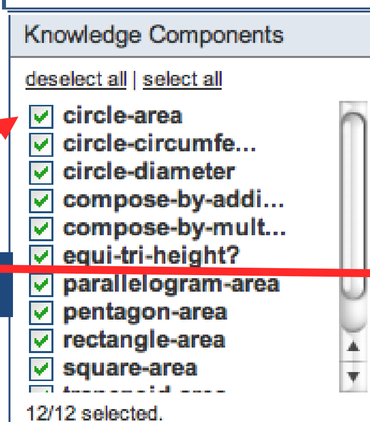
DataShop's learning curve tools aid discovery of better cognitive models





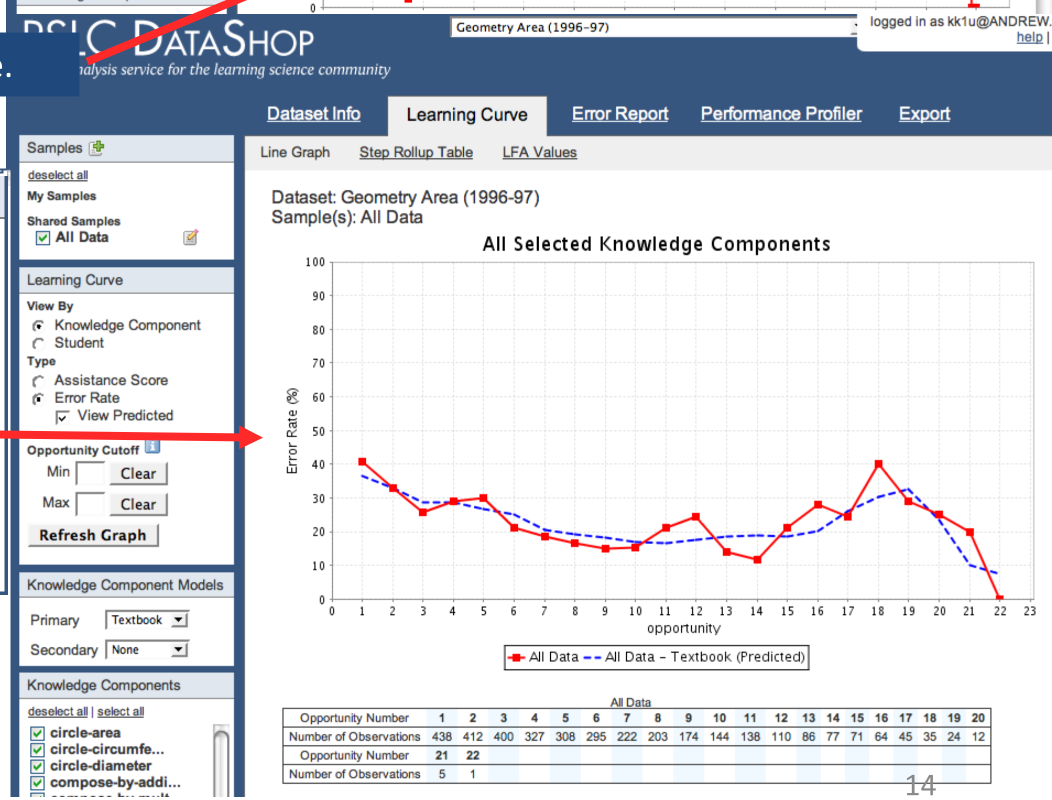
Without decomposition, using just a single "Geometry" KC,

no smooth learning curve.



But with decomposition, 12 KCs for area concepts,

a smoother learning curve.



Upshot: Cognitive models of student learning can be discovered from data

DataShop's "leaderboard" ranks discovered cognitive models

100s of datasets coming from ed tech in math, science, & language

PSLC DataShop

a data analysis service for the learning science community

Geometry Area (1996-97)

KC Models

Export | Import

Sort by Cross Validation RMSE (item stratified) ascending

	Model Name	KCs	Observations with KCs	AIC	BIC	Cross Validation*: RMSE (student stratified)	Cross Validation*: RMSE (item stratified)	Cross Validation*: RMSE (unstratified)	Cross Validation*: Observations (unstratified)	
	LFASearchAICWholeModel3	18	5104	4989.31	5610.39	0.407966	0.397185	0.396639	5104	
	LFASearchAICWholeModel2	17	5104	4992.65	5600.67	0.410169	0.397783	0.398564	5104	
	LFASearchAICWholeModel0	17	5104	5001.46	5609.47	0.408579	0.397816	0.39889	5104	
	LFASearchModel1.context-single	13							5104	
	LFASearchModel1-renamed	11							5104	
	LFASearchModel1	11							5104	
	LFASearchAICWholeModel1	16							5104	
	LFASearchModel1-backward	13	5104	5035.28	5590.99	0.410264	0.400649	0.399619	5104	
	xDecmpTrapCheat	12	5104	5063.6	5606.24	0.409086	0.401162	0.402429	5104	
	LFASearchModel1-renamed&chgd.2	11	5104	5042.21	5571.77	0.408617	0.4013	0.401477	5104	
	LFASearchModel1-renamed&chgd	11	5104	5065.98	5595.54	0.409898	0.401302	0.400725	5104	
	LFASearchModel1-context	12	5104	5030.46	5573.1	0.40916	0.401557	0.399431	5104	
	LFASearchModel1-back-context	13	5104	5033.83	5589.54	0.410125	0.401796	0.401784	5104	
	LFASearchModel0	10							5104	
	DecompArithDiam	13							5104	
	Unique-step	132	5083	5071.87	7182.25	0.413822	0.402627	0.400798	5082	
	textbook2	13	5104	5106.1	5661.82	0.412562	0.402893	0.402177	5104	
	DecomposeArith	12	5104	5085.89	5628.53	0.411786	0.402956	0.401513	5104	

Some models are *machine* generated (based on human-generated learning factors)

Some models are *human* generated

Statistical Model of Learning

Additive Factors Model

Correct?

Step j uses skill k

$$\log \frac{p_{ij}}{1-p_{ij}} = \theta_i + \sum_k \beta_k Q_{kj} + \sum_k Q_{kj}(\gamma_k T_{ik})$$

Opp count

GIVEN:

p_{ij} = probability student *i* gets step *j* correct

Q_{kj} = each knowledge component *k* needed for this step *j*

T_{ik} = opportunities student *i* has had to practice *k*

ESTIMATED:

θ_i = proficiency of student *i*

β_k = difficulty of KC *k*

γ_k = gain for each practice opportunity on KC k

Cen, Koedinger, & Junker (2006)
Draney, Pirolli, & Wilson (1995)
Spada & McGaw (1985)

Metrics for model prediction

PSLC DataShop
a data analysis service for the learning science community

Model Name	KCs	Observations with KCs	AIC	BIC	Cross Validation*			
					RMSE (student stratified)	RMSE (item stratified)	RMSE (unstratified)	Observations (unstratified)
LFASearchAICWholeModel3	18	5104	4989.31	5610.39	0.407966	0.397185	0.396639	5104
LFASearchAICWholeModel2	17	5104	4992.65	5600.67	0.410169	0.397783	0.398564	5104
LFASearchAICWholeModel0	17	5104	5001.46	5609.47	0.408579	0.397816	0.39889	5104
LFASearchModel1.context-single	19	5104	4990.76	5624.93	0.409601	0.398209	0.399035	5104

- AIC & BIC penalize for more parameters, fast & consistent
- Cross Validation, 10 fold
  - ➢ Target root mean squared error (RMSE)
  - ➢ *Stratified* by student, item, or not
  - ➢ Student models predict future performance
    => item stratified is best choice

Automated search for better models

Learning Factors Analysis (LFA)

(Cen, Koedinger, & Junker, 2006)

- Method for discovering & evaluating cognitive models
- Finds Q matrix that best predicts learning
- Inputs
  - Data: Student success on tasks over time
  - P matrix: Factors hypothesized to explain learning
  - Initial Q matrix
- Outputs
  - Rank order of most predictive Q matrix
  - Parameter estimates for each

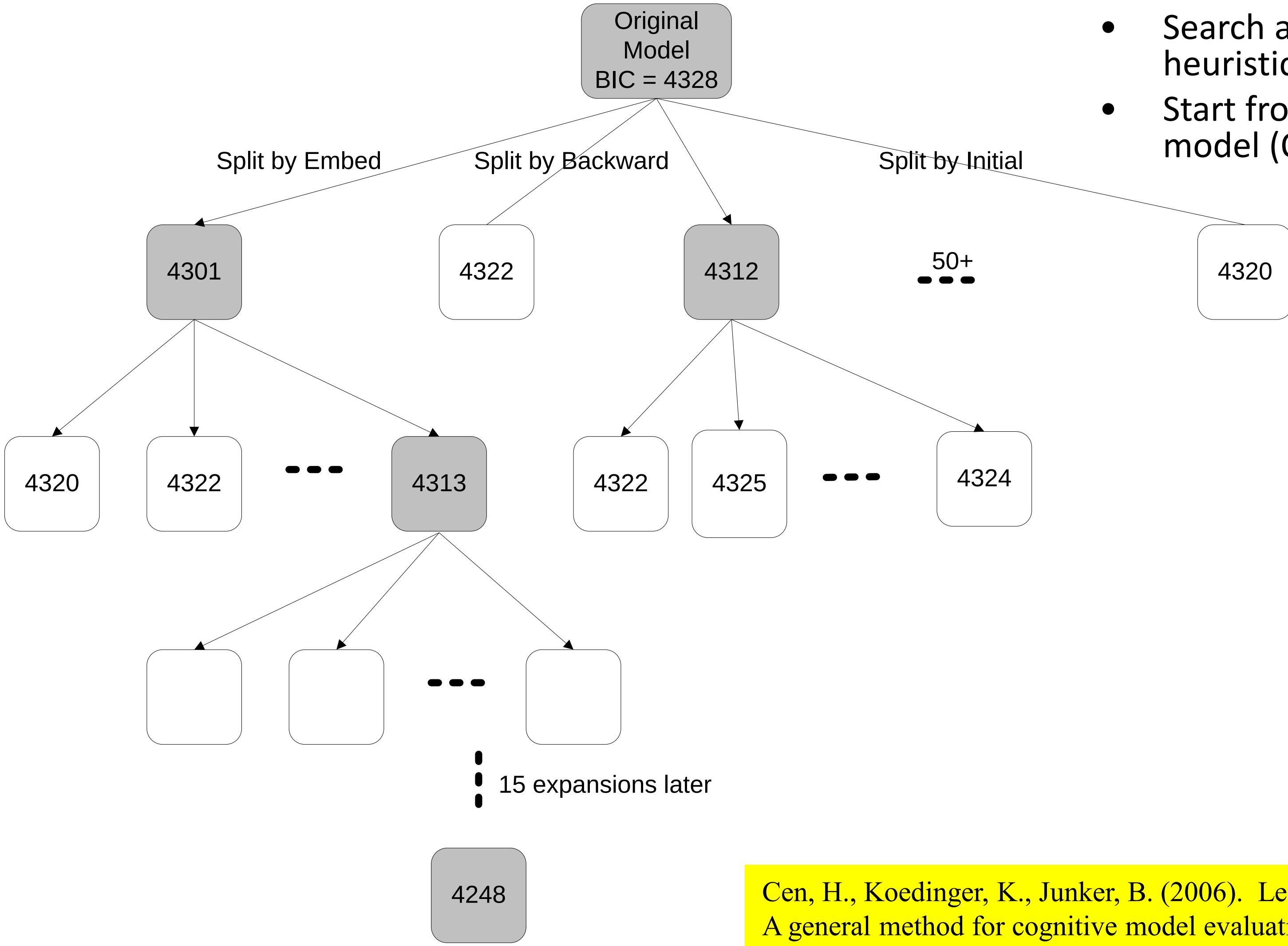
Simple search process example: modifying Q matrix by factor in P matrix to get new Q' matrix

Q

Problem Step	Mult	Sub
2*8-30 => 16-30	1	0
16 - 30 => -14	0	1
30-2*8 => 30-16	1	0
30-16 => 14	0	1

- Q matrix factor Sub *split* by P matrix factor Neg-result
- Produces new Q matrix
- Two new KCs (Sub-Pos & Sub-Neg) replace old KC (Sub)
  - Redo opportunity counts

LFA: Best First Search Process



- Search algorithm guided by a heuristic: BIC
- Start from an existing cog model (Q matrix)

Cen, H., Koedinger, K., Junker, B. (2006). Learning Factors Analysis: A general method for cognitive model evaluation and improvement. *8th International Conference on Intelligent Tutoring Systems*.

But where does P matrix come from?

Frequent questions about LFA:

- Isn't all the feature labeling too much work?
- Wouldn't it be better to have a data mining algorithm that doesn't require it?

Answers: No and no!

- Better prediction without interpretation does not help to redesign instruction!

=> Extract labels from "normal" course of work …

Scientist "crowd"sourcing: Feature input (P) comes "for free"

PSLC DataShop

a data analysis service for the learning science community

Geometry Area (1996-97)

Export | Import

Sort by Cross Validation RMSE (item stratified) ascending

	Model Name	KCs	Observations with KCs	AIC	BIC	Cross Validation* RMSE (student stratified)	Cross Validation* RMSE (item stratified)	Cross Validation* RMSE (unstratified)	Cross Validation* Observations (unstratified)	
	LFASearchAICWholeModel3	18	5104	4989.31	5610.39	0.407966	0.397185	0.396639	5104	
	LFASearchAICWholeModel2	17	5104	4992.65	5600.67	0.410169	0.397783	0.398564	5104	
	LFASearchAICWholeModel0	17	5104	5001.46	5609.47	0.408579	0.397816	0.39889	5104	
	LFASearchModel1.context-single	19	5104	4990.76	5624.93	0.409601	0.398209	0.399035	5104	
	LFASearchModel1-renamed	11	5104	[illegible]	[illegible]	[illegible]	[illegible]	0.39813	5104	
	LFASearchModel1	11	5104	[illegible]	[illegible]	[illegible]	[illegible]	0.399224	5104	
	LFASearchAICWholeModel1	16	5104	[illegible]	[illegible]	[illegible]	[illegible]	0.399845	5104	
	LFASearchModel1-backward	13	5104	[illegible]	[illegible]	[illegible]	[illegible]	0.399619	5104	
	xDecmpTrapCheat	12	5104	5063.6	5606.24	0.409086	0.401162	0.402429	5104	
	LFASearchModel1-renamed&chgd.2	11	5104	5042.21	5571.77	0.408617	0.4013	0.401477	5104	
	LFASearchModel1-renamed&chgd	11	5104	5065.98	5595.54	0.409898	0.401302	0.400725	5104	
	LFASearchModel1-context	12	5104	5030.46	5573.1	0.40916	0.401557	0.399431	5104	
	LFASearchModel1-back-context	13	5104	5033.83	5589.54	0.410125	0.401796	0.401784	5104	
	LFASearchModel0	10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5104	
	DecompArithDiam	13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5104	
	Unique-step	132	5083	5071.87	7182.25	0.413822	0.402627	0.400798	5082	
	textbook2	13	5104	5106.1	5661.82	0.412562	0.402893	0.402177	5104	
	DecomposeArith	12	5104	5085.89	5628.53	0.411786	0.402956	0.401513	5104	

Union of all hypothesized KCs in human generated models

Some models are *human* generated

Does it work?

Koedinger, McLaughlin, & Stamper (2012). Automated student model improvement. In *Proceedings of the Fifth International Conference on Educational Data Mining*. [Conference best paper.]

Applying LFA across domains

Dataset	Content area	RMSE			Median Learning slope (logit)	
		Orig in-use	Best-hand	Best-LFA	Best-hand	Best-LFA
Geometry 9697	Geometry area	0.4129	0.4033	0.4011	0.07	0.11
Hampton 0506	Geometry area	NA	0.4022	0.4012	0.03	0.04
Cog Discovery	Geometry area	NA	0.3250	0.3244	0.16	0.16
DFA-318	Story problems	0.4461	0.4407	0.4405	0.07	0.17
DFA-318-main	Story problems	0.4376	0.4287	0.4266	0.09	0.17
Digital game	Fractions	0.4442	0.4396	0.4346	0.17	0.14
Self-explanation	Equation solving	NA	0.4014	0.3927	0.01	0.04
IWT 1	English articles	0.4262	0.4110	0.4068	0.10	0.12
IWT 2	English articles	0.3854	0.3854*	0.3806	0.12	0.16
IWT 3	English articles	0.3970	0.3965	0.3903	0.05	0.15
Statistics-Fall09	Statistics	0.3648	0.3527	0.3353	**	0.09

Variety of domains & technologies

11 of 11 improved models

9 of 11 equal or greater learning

Interpreting student model improvements

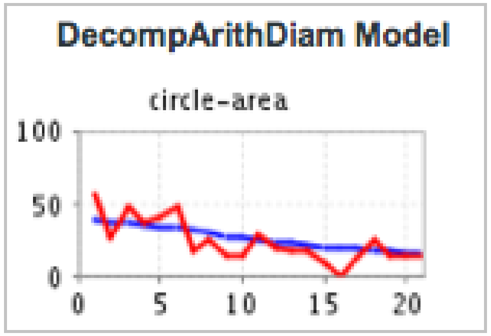
Method

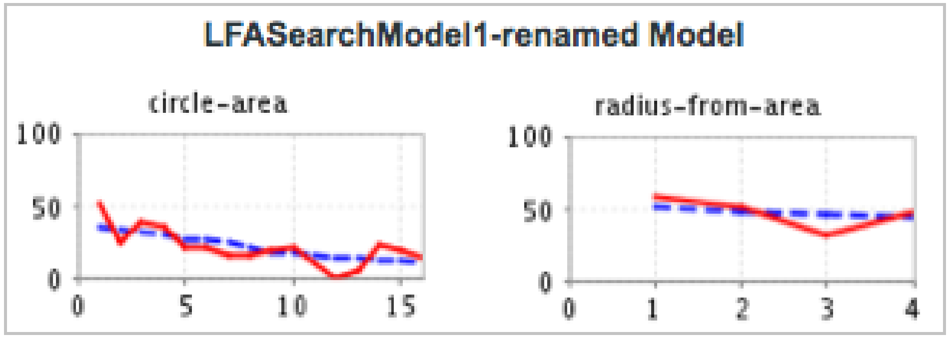
- Isolate improvement in KCs from base model to new models by computing reduction in RMSE

Example analysis from Geometry 9697 dataset

Original model KCs	% reduction in RMSE		
	orig->hand	hand->LFA	orig-LFA
CIRCLE-RADIUS	5.8%	4.0%	9.5%
COMPOSE-BY-ADDITION	5.2%	0.3%	5.5%
TRIANGLE-SIDE	10.0%	1.2%	11.1%
Range of the 12 other KCs	-.5 to 3.4%	-.3 to 1.0%	-.2 to 3.1%

New KC revealed from LFA search





KC Values For DecompArithDiam Model

KC Name	Intercept (logit)	Intercept (probability)	Slope
circle-area	0.31	0.58	0.068

Greater slope indicates better model of learning and transfer

Uses forward strategy

KC Values For LFASearchModel1-renamed Model

KC Name	Intercept (logit)	Intercept (probability)	Slope
circle-area	0.44	0.61	0.106
radius-from-area	-0.38	0.41	0.165

Uses backward strategy

Working backgrounds is not a general skill (in this domain)

Circle-area backward (54%) harder than forward (80%)

But no backward vs. forward diff in general

pentagon area: 66% vs. 62%

trapezoid area: 54% vs. 55%

…

Devil's in the details

Interpretation:

- Unique feature of circle-area backwards is need for a *square root operation*
- More practice & instruction needed

Implications for tutor redesign

Recommendations:

- More focus on problems requiring square root
  - More area-to-radius & square-area-to-side problems
- Merge all *other* forward/backward distinctions
- Modify skill bars

Expected Outcomes:

- Reduce time to mastery for other area formulas
- More time for practice on square root

Conclusions

- Combine human & machine intelligence
  - Student crowdsourcing: data from widely fielded systems
  - Scientist crowdsourcing (in the small): Use leaderboard to gain human participation in feature labeling
  - Results in better prediction *AND interpretable and useful models*

- Better predictions across 11 datasets & a variety of domains
  - Models overlap => small prediction error reduction, but …
- Isolated improvements are large and useful
  - Yield tutor redesign recommendations
=> more efficient & effective learning & transfer

Comment on the Zeitgeist…

- Lots of great excitement, new ideas & smart people coming to the educational technology table
- Keep it coming!
- Find partners & do not get discouraged as it gets tough
  - Education is a **hugely** complex system
  - Complex systems have many failure joints
  - Devil's in the details

- Complexity/opportunity in instruction options too:

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

Thank you

- Crowdsourcing Cognitive Models for Assessment, Tutoring, and In-Game Support

  Cognitive Tutors personalize online education by adapting problem-based instruction both within problems, by providing feedback and instruction specific to students' solution progress, and between problems, by selecting and pacing activities based on students' performance history. The approach relies on a cognitive model of the unobservable mental skills and concepts needed for success across problem-solving tasks. Widespread use of tutors (and educational games and simulations too) is producing vast data that can be used to create and improve cognitive models to drive better adaptive instruction. LearnLab's DataShop (http://learnlab.org/datashop) is the world's largest open repository of such data. I will discuss educational data mining algorithms to create, test, and improve cognitive models for better personalized assessment and tutoring within educational technologies. Large crowds of students and, more recently, small crowds of learning researchers are being leveraged in these approaches. In a recent study, we employed our Learning Factors Analysis algorithm on eleven different DataShop datasets to automatically discover better cognitive models in all cases.

Extra slides

Sample of log data used to generate learning curves

Items = steps in tasks with feedback

Q-matrix in single column: works for single skill items

Opportunities student has had to learn skill

Student	Step (Item)	Skill (KC)	Opportunity	Success
A	p1s1	Circle-area	1	0
A	p2s1	Circle-area	2	1
A	p2s2	Rectangle-area	1	1
A	p2s3	Compose-by-addition	1	0
A	p3s1	Circle-area	3	0

Example in Geometry of "split" based on factor in P matrix

Original Q matrix

Factor in P matrix

Student	Step	Skill	Opportunity	Embed
A	p1s1	Circle-area	1	alone
A	p2s1	Circle-area	2	embed
A	p2s2	Rectangle-area	1	
A	p2s3	Compose-by-add	1	
A	p3s1	Circle-area	3	alone

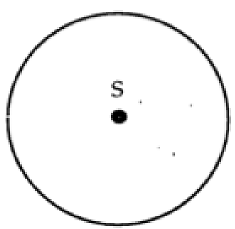
After Splitting Circle-area by *Embed*

New Q matrix

Revised Opportunity

Student	Step	Skill	Opportunity
A	p1s1	Circle-area-alone	1
A	p2s1	Circle-area-embed	1
A	p2s2	Rectangle-area	1
A	p2s3	Compose-by-add	1
A	p3s1	Circle-area-alone	2

LAWN_SPRINKLER: SECTION FOUR, #2



Problem Statement

If a lawn sprinkler spins 360 degrees and has a 17 ft spray, find the area of lawn that will be watered if the sprinkler is not moved.

	Length of Spray (OD) - Radius	Watered Lawn - Circle
Units	feet	sq. feet
Question 1	17	907.92

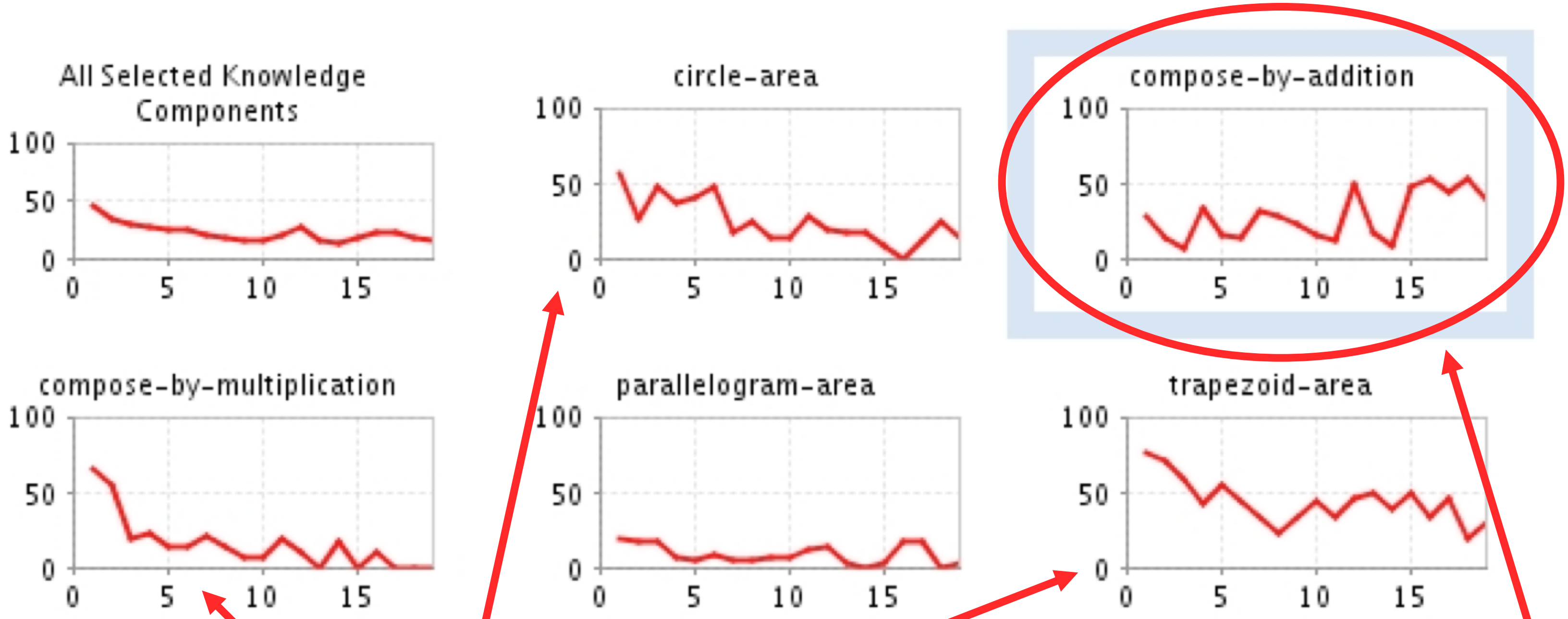
LAWN_SPRINKLER_2: SECTION FOUR, #5

Problem Statement

A lawn sprinkler spins 360 degrees and has a 1 ft spray. If the rectangular lawn is 5 feet by 9 feet, find the area of lawn that will not be watered if the sprinkler is not moved.

	Length of Spray (OD) - (Radius)	Watered Lawn (Circle)	Area of Total Lawn (Rectangle)	Area of Unwatered Lawn
Units	cm	cm	sq. cm	cm
Question 1	1	3.1416	45	41.86

DataShop visualizations help identify potentially bad KCs

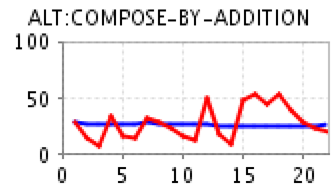


Many curves show a reasonable decline

Some do not => Opportunity to improve model!

Example model modification with implications for tutor redesign

KC Values For Original Model

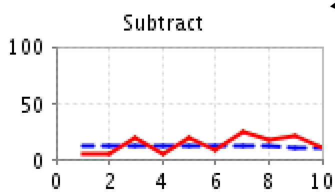


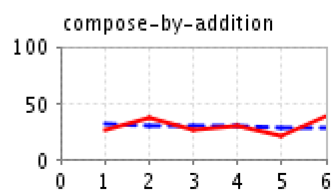
non-smooth, not low, not declining

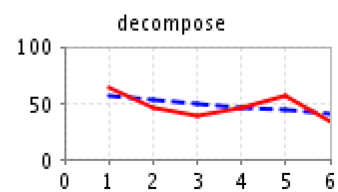
KC Name	Intercept (logit)	Intercept (probability)	Slope
ALT:COMPOSE-BY-ADDITION	1.04	0.74	0

KC Values For DecompArithDiam Model

Original Model KC split into 3 KCs

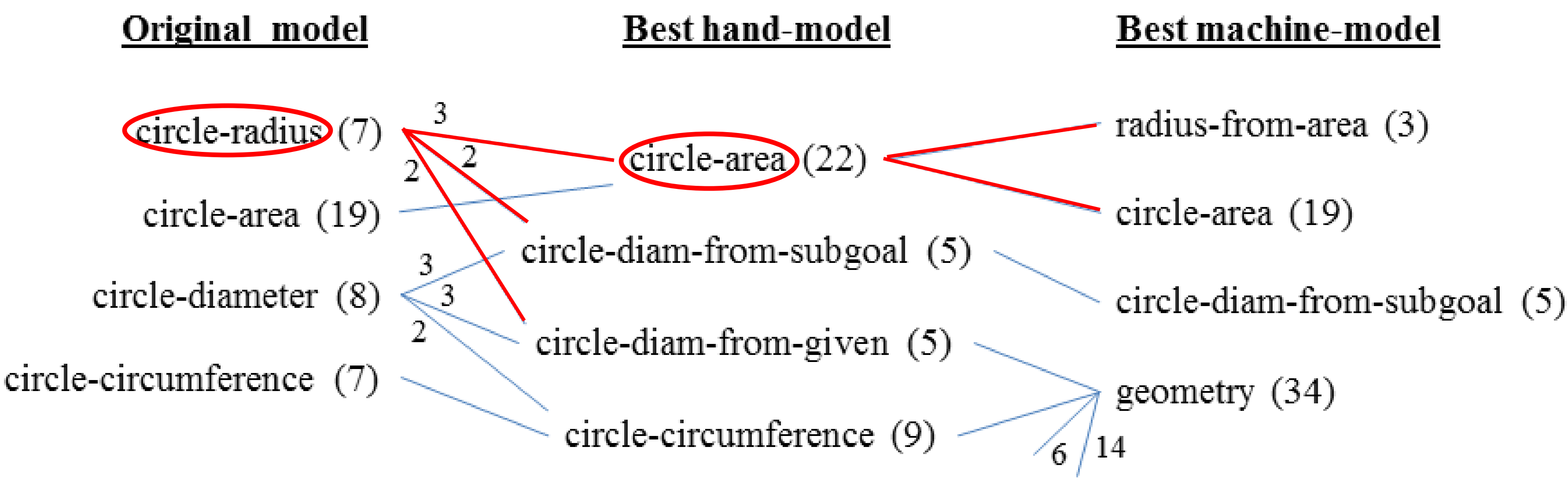






KC Name	Intercept (logit)	Intercept (probability)	Slope
Subtract	2.05	0.89	0
compose-by-addition	0.84	0.7	0
decompose	-0.56	0.36	0.148

Splitting and combining of circle-radius and other related hypothesized KCs



- Circle-radius splits to circle-area, circle-diam-from-subgoal, and circle-diam-from-given
- Target skill for all 3 is computing radius
- Circle-area further splits to circle-area and radius-from-area

Future Work

- Variations on Learning Factors Analysis
  - Use add operator in search (models partial transfer)
  - Try other statistical models: BKT, PFA …
  - Other search methods: beam search, new heuristics …
- In vivo experiments to test model discoveries
  - compare learning with revised vs. original ed tech
- Apply human-machine discovery approach in other domains
  - Humans do feature engineering; machine uses
  - Humans & machine compete & cooperate through leaderboard