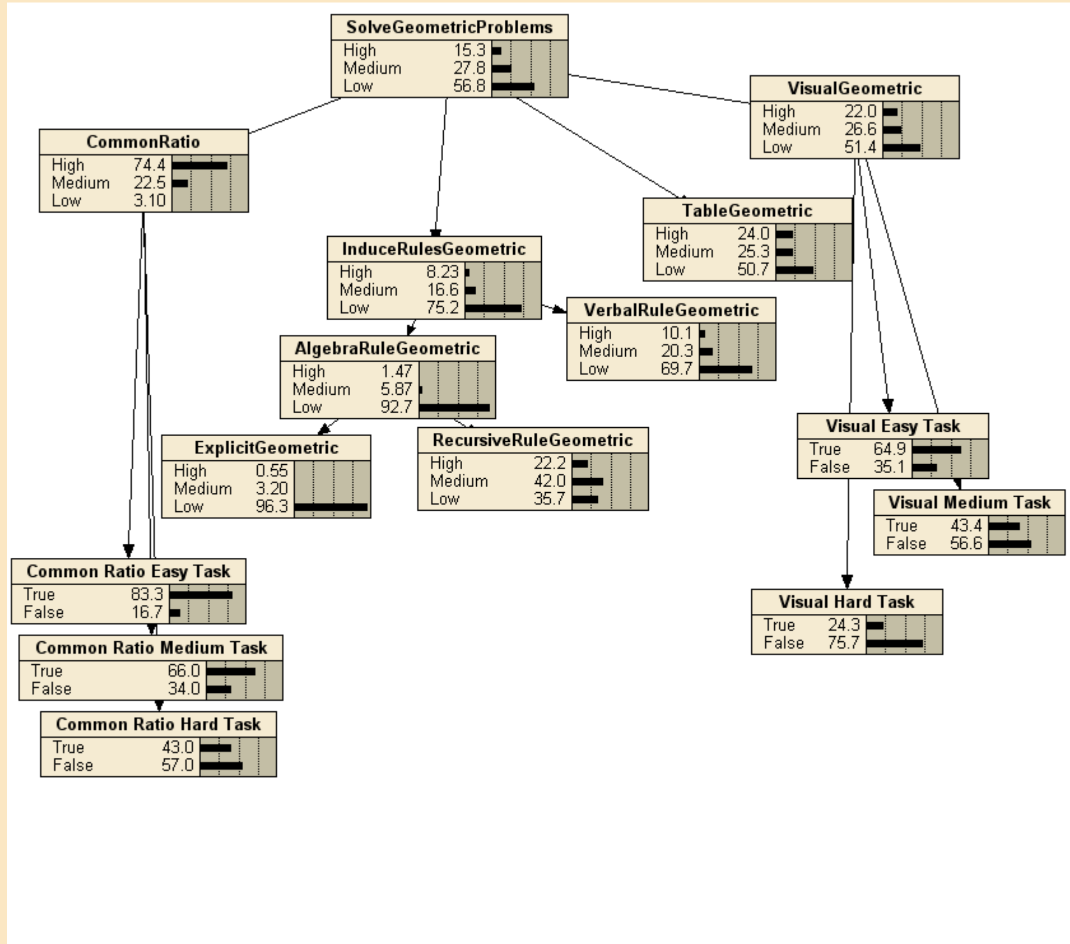
ACED: An Example of Using Bayesian Networks for Cognitive Diagnosis

Russell Almond
Val Shute
Florida State University
College of Education
Educational Psychology and Learning Systems
ralmond@fsu.edu

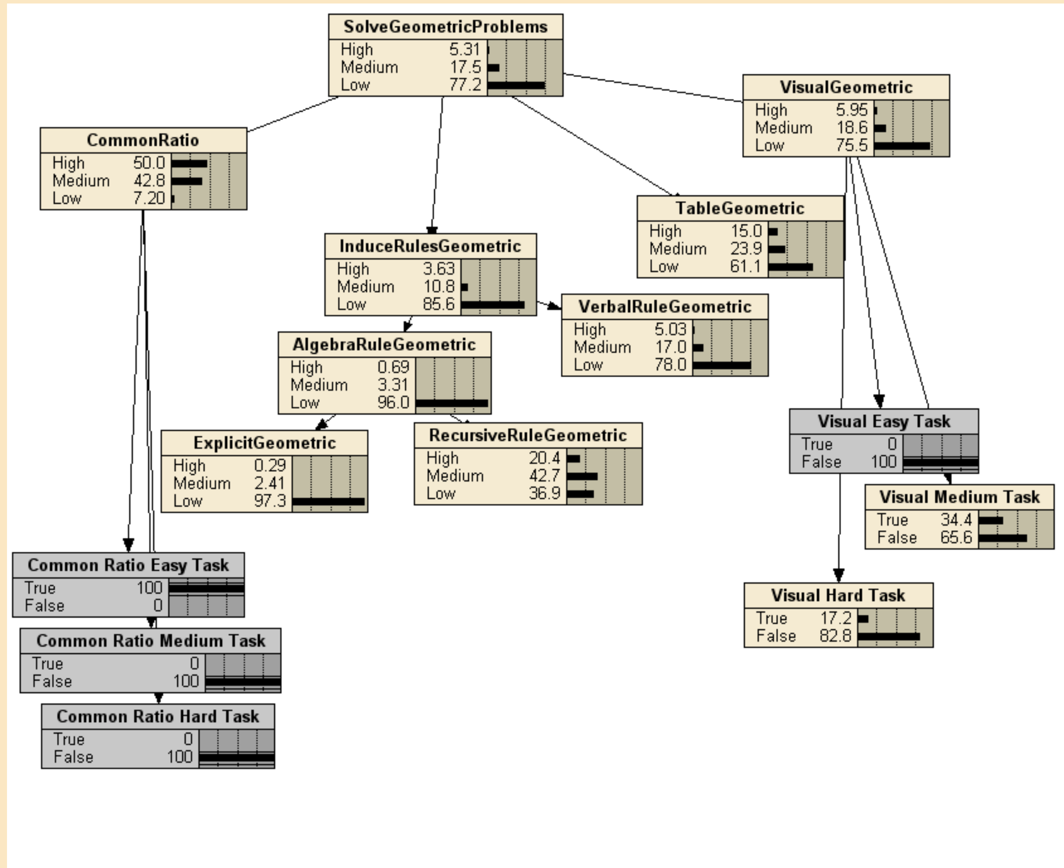
Bayesian Networks

- Use acyclic directed graph to represent factorization of complex multivariate probability distribution.
- Separation in graph implies conditional independence.
- Fast computational algorithms
- Discrete Bayesian networks supported by a number of software tools (Netica http://www.norsys.com/)



Real-time Inference

- Entering Evidence about state of observable variables instantly updates probabilities for unobserved variables
- Inferences based on direct and indirect evidence
- Calculate expected weight of unobserved evidence
- Think about one student at a time
- Scoring engine embedded in a learning system



ACED Background

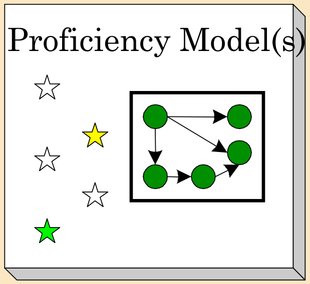
- ACED (Adaptive Content with Evidence-based Diagnosis)
- Val Shute (PD), Aurora Graf, Jody Underwood, Eric Hansen, Peggy Redman, Russell Almond, Larry Casey, Waverly Hester, Steve Landau, Diego Zapata
- Domain: Middle School Math, Sequences
- Project Goals:
  - Adaptive Task Selection
  - Diagnostic Feedback
  - Accessibility
- Bayesian Network Inference Engine (StatShop)
  - EWOE-based task selection option

Evidence Centered Design

- The Evidence Centered Design process is a series of procedures which center around the questions:
  - “What can we observe about an examinee's performance which will provide evidence that the examinee has or does not have the knowledge, skills and abilities we wish to make claims about?”
  - “How can we structure situations to be able to make those observations?”
- This process results in a formal design for an assessment we call the **Conceptual Assessment Framework (CAF)**

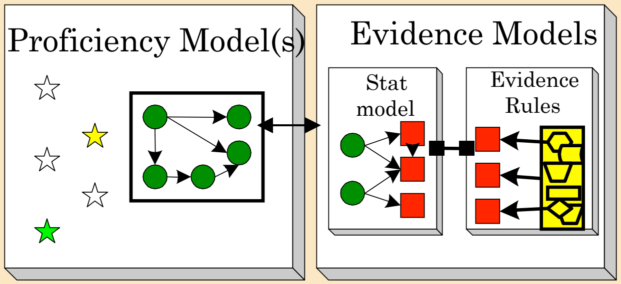
Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model



Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model
How we measure = **Evidence** Model

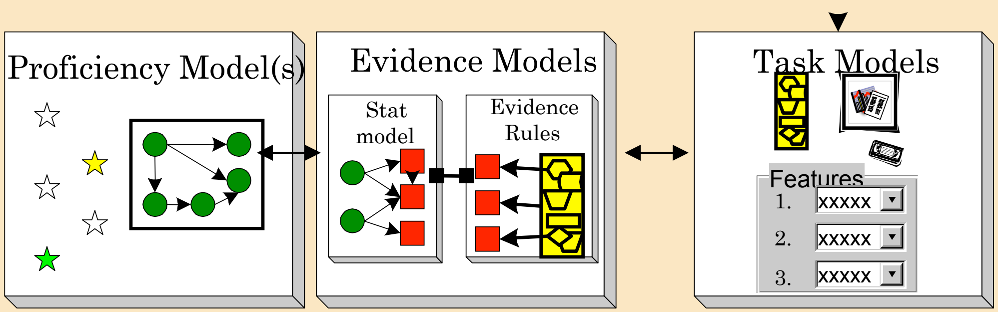


Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model

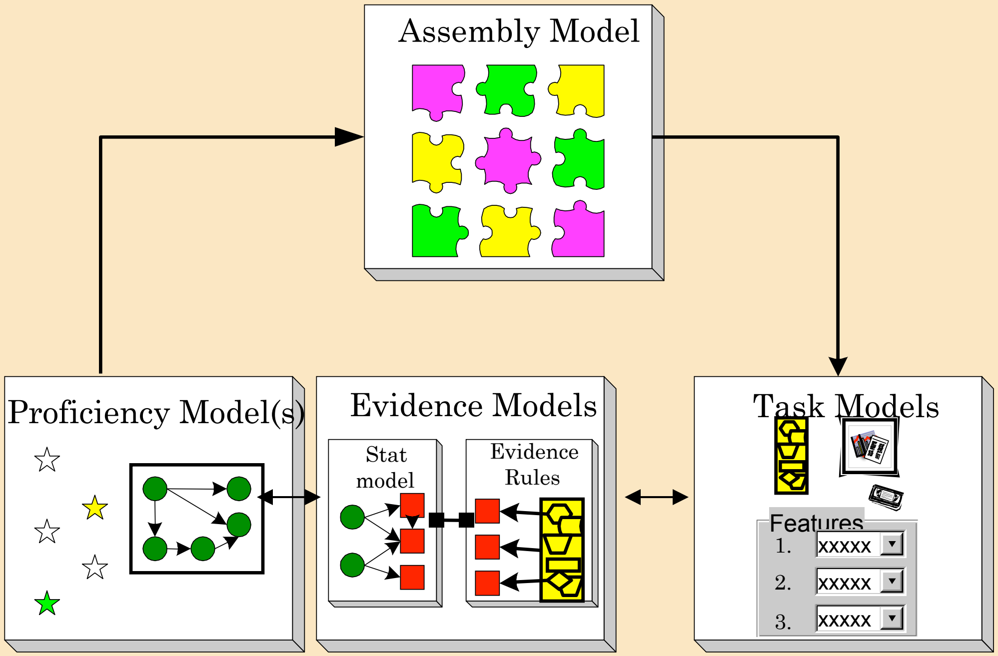
How we measure = **Evidence** Model

Where we measure = **Task** Model



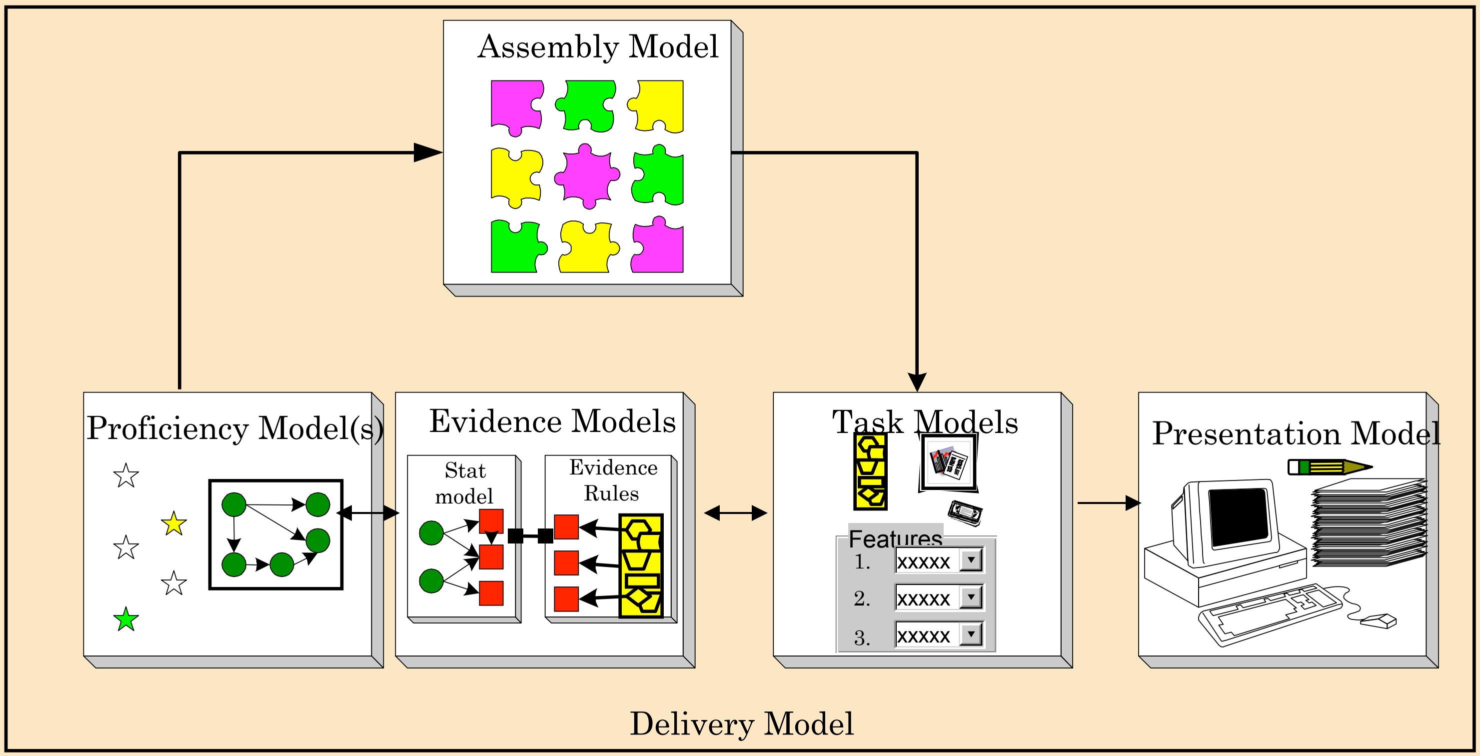
Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model
How we measure = **Evidence** Model
Where we measure = **Task** Model
How Much we measure = **Assembly** Model

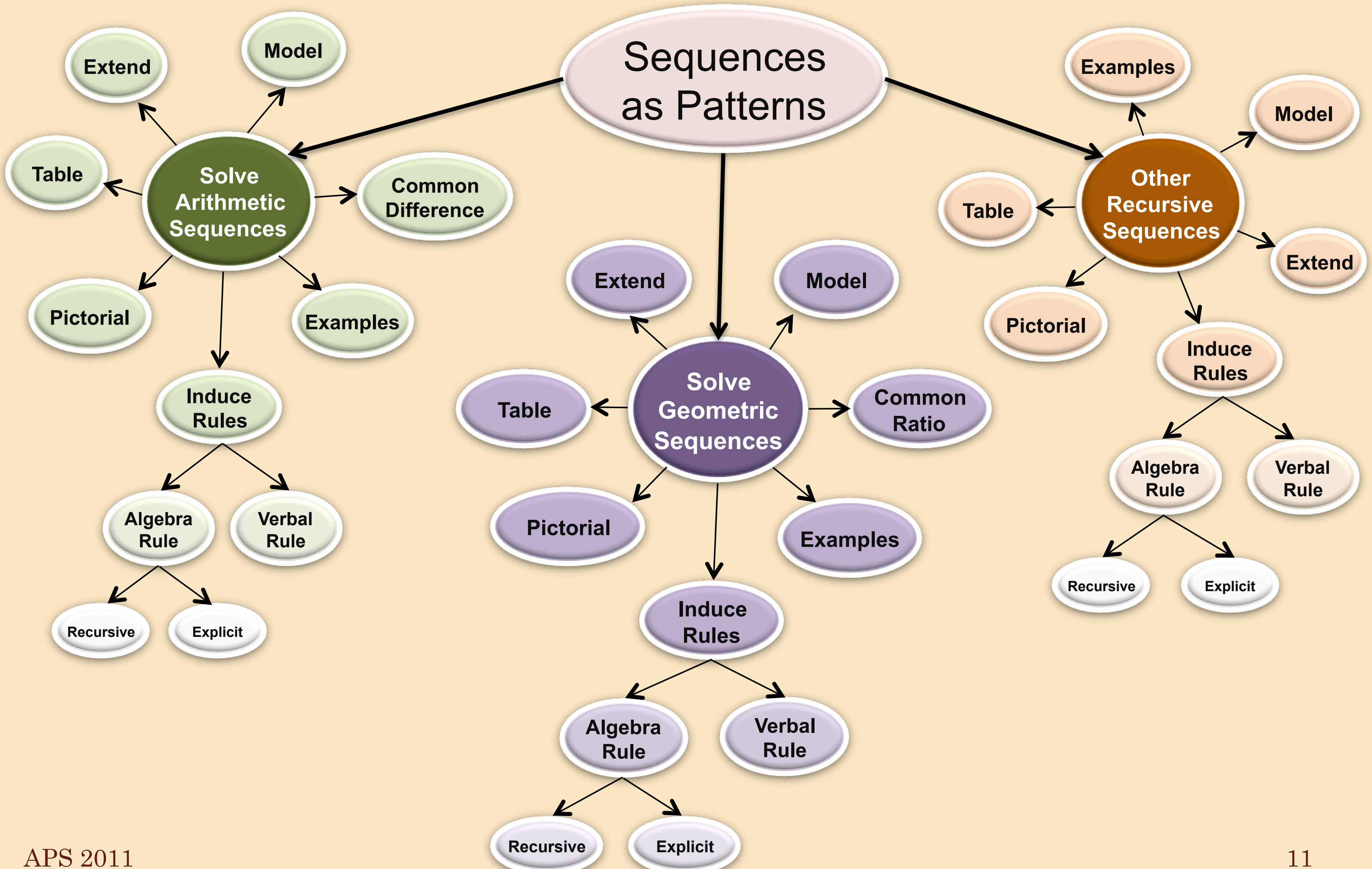


Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model
How we measure = **Evidence** Model
Where we measure = **Task** Model
How Much we measure = **Assembly** Model
Customization = **Presentation & Delivery** Models



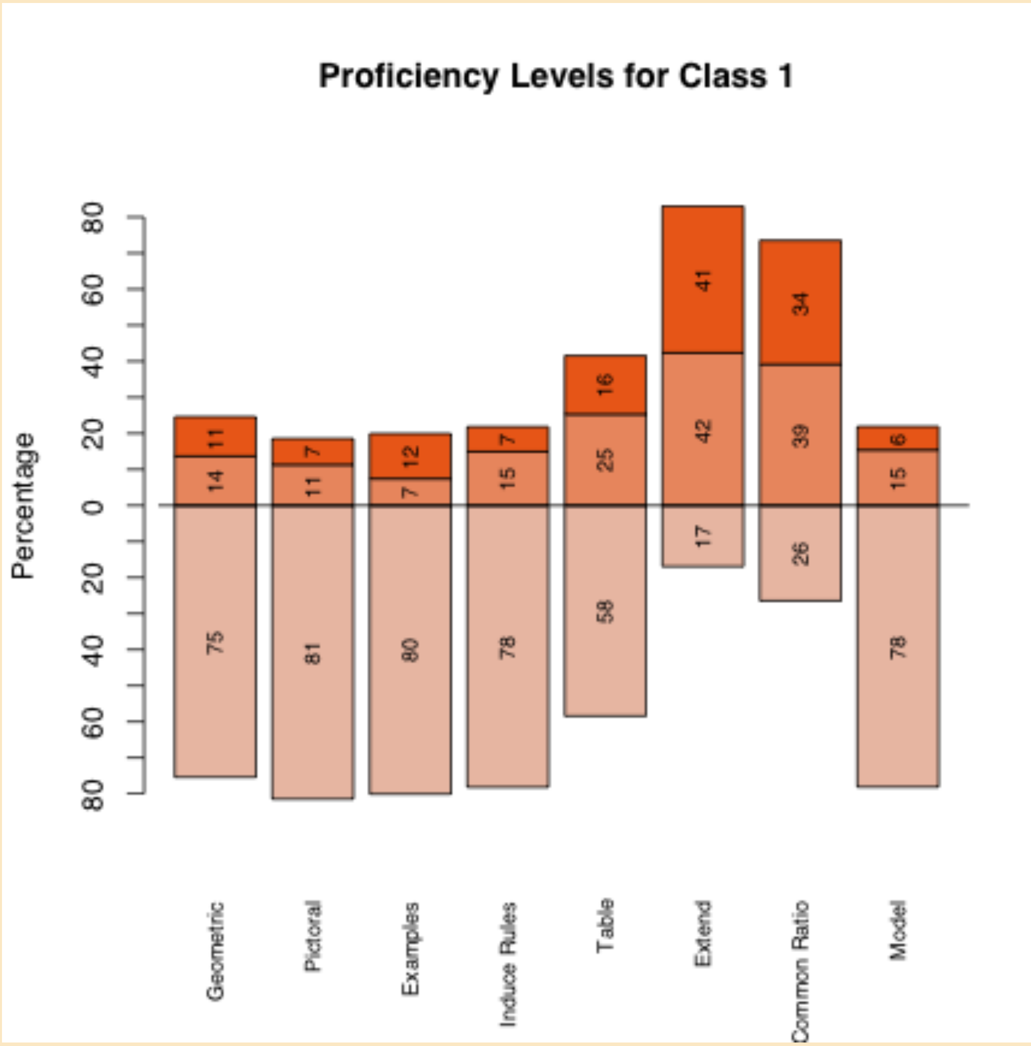
ACED Proficiency Model



Proficiency Correlation Model

- Represents distribution of skills in population distribution
- Subject matter expert provided three numbers for each proficiency variable:
  - Correlation with parent(s)
  - Prevalence (intercept term)
  - Confidence (learned in using parameters)
- Psychometricians converted into numbers
- Could use inverse correlation matrix to determine graphical structure (not used in ACED)

ACED Scores



- For Each Proficiency Variable
  - Marginal Distribution
  - Modal Classification
  - EAP Score (High=1, Low=-1)

ACED Task Models

- General Task Models used to construct specific Tasks
- Task Models designed to target various proficiency variables
- Task Model Variables are parameters of the Task Model that can be manipulated to produce task variants
    - Equivalent difficulty
    - Varying difficulty (ACED build “Easy”, “Medium” and “Hard” variants)
    - Differing evidentiary focus (targeting different proficiency variables)

Typical Task

Default Frameset - Microsoft Internet Explorer provided by ETS

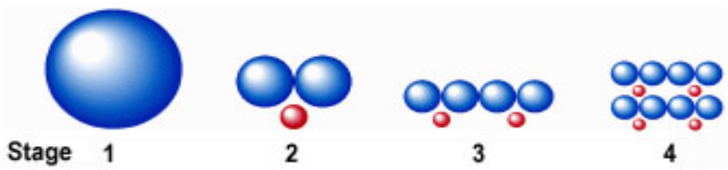
99 MINUTES

ADAPTIVE E-LEARNING

QUESTION 1 OF 1 S1

Sequence	Blue	Red	Total
1	1	0	1
2	2	1	3
3	4	2	6
4	8	4	12
...	...	...	...
N	**A**	**B**	**C**

Katie is a biochemist. During her last trip to the Amazon rainforest, she brought back some leaves from an exotic plant. She extracted a substance from those leaves that had some amazing properties. One drop of the substance on a given cell produced a doubling of the cell, along with a smaller bonus cell (see Stages 1 and 2, below). The same pattern was found in consecutive trials (see Stages 3-4).



She made a table of her findings. Your task is to figure out how many blue, red, and total cells would be present in the 8th sequence. Complete the table by filling in the values for A, B, and C (where N = 8).

Enter the value for **A:**

Enter the value for **B:**

Enter the value for **C:**

CALC PREVIOUS REVIEW NEXT

Informative Feedback

Default Frameset - Microsoft Internet Explorer provided by ETS

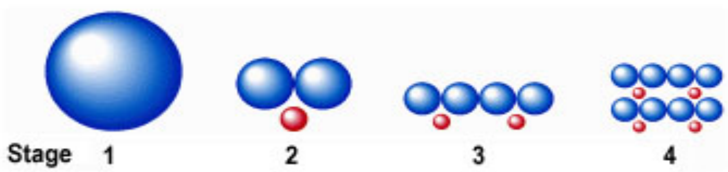
99 MINUTES

ADAPTIVE E-LEARNING

QUESTION 1 OF 1 S1

Sequence	Blue	Red	Total
1	1	0	1
2	2	1	3
3	4	2	6
4	8	4	12
...	...	...	...
N	**A**	**B**	**C**

Katie is a biochemist. During her last trip to the Amazon rainforest, she brought back some leaves from an exotic plant. She extracted a substance from those leaves that had some amazing properties. One drop of the substance on a given cell produced a doubling of the cell, along with a smaller bonus cell (see Stages 1 and 2, below). The same pattern was found in consecutive trials (see Stages 3-4).



She made a table of her findings. Your task is to figure out how many blue, red, and total cells would be present in the 8th sequence. Complete the table by filling in the values for A, B, and C (where N = 8).

Enter the value for **A:**

Enter the value for **B:**

Enter the value for **C:**

Sorry, that's not right. Make sure you fill in the row for $N = 8$. It looks like you figured out the right information for Stage 5. The correct answers are: $A = 128$, $B = 64$, and $C = 192$ (the sum of $A + B$).

CALC PREVIOUS REVIEW NEXT

Evidence Models

- Evidence models link task work products to proficiency variables
- Evidence Rules define *observable outcome variables* from raw response
  - Automatic scoring of numeric & graphical responses
  - Determines appropriate feedback
- Statistical evidence model defines probability of observed outcome given proficiency profile

DiBello's Quasi-IRT Models

- Use IRT models to produce conditional probability tables for Bayes nets
- For each state, $s_{k,m}$, of parent variable, S_k assign an *effective theta* $\tilde{\theta}_{k,m}$
- Use IRT 2PL model to assign probabilities

$$P(X_i = 1 | S_{j,1}, \ldots, S_{j,K}) =$$
$$\text{logit}^{-1}\left[g\left(a_{i,1}\tilde{\theta}_{1,S_{j,1}}, \ldots, a_{i,K}\tilde{\theta}_{1,S_{j,K}}\right)\right] - b_j$$

- The $a_{i,k}$ parameters are called *discriminations*
- The b_i parameter is called the *difficulty*

Structure Function

- When there are multiple parents
- Combine Using a Structure Function

$$g\left(a_{i,1}\tilde{\theta}_{1,S_{j,1}}, \ldots, a_{i,K}\tilde{\theta}_{1,S_{j,K}}\right)$$

- Possible Structure Functions:
  - Compensatory = average
  - Conjunctive = min
  - Disjunctive = max
  - Inhibitor = $\begin{cases} \tilde{\theta}_{2,S_2} & \text{if } \tilde{\theta}_{1,S_1} > \theta^* \\ \tilde{\theta}_{2,0} & \text{otherwise} \end{cases}$
- Others possible

ACED: EM level Q-Matrix

TaskModel	common	distinguish	examples	explicit	extend	model	recursive	table	verbal	visual	*Anchor*	Tasks
commonRatioTask	+										*commonRatio*	1a,b, 2a,b, 3a,b
examplesGeometricTask			+								*examples*	1a,b, 2a,b, 3a,b
explicitGeometricTask				+							*explicit*	1a,b, 2a,b, 3a,b
extendGeometricTask					+						*extend*	1a,c, 2a,c, 3a,b
solveGeometricTask		+									*distinguish*	1a,b, 2a,b, 3a,b
modelExtendTableGeometricTask					+	++		+			*model*	1a,b, 2a,b, 3a,b
recursiveRuleGeometricTask							+				*recursive*	1a,b, 2a,b, 3a,b
tableExtendGeometricTask					+			++			*table*	1a,b, 2a,b, 3a,b
verbalRuleExplicitModelGeometricTask				+		+			++		*verbal*	2a, 3a
verbalRuleExtendModelGeometricTask					+	+			++		*verbal*	1a,b, 2a
verbalRuleModelGeometricTask						+			++		*verbal*	3a
visualExplicitVerbalRuleModelGeometric				+		+			+	++	*visual*	3a, 3b
visualExtendGeometricTask					+					++	*visual*	1a,b,c, 2a, 3a
visualExtendTableModelVerbalRuleGeom					+	+		+	+	++	*visual*	2a
visualExtendVerbalRuleModelGeometric					+	+			+	++	*visual*	2a

- Can do simple design tests at this level
- Aliasing: For each proficiency, need either
  - Simple structure task (taps only one skill)
  - Block triangular structure -- a defining task for each proficiency which taps only that skill and other previously designed skills
- Note: Difficulty level coded in task names (1,2,3)

ACED Assembly Model

- Controls composition of test form
- ACED had two modes:
  - Linear: Fixed sequence
  - Adaptive
    - Pick task that provided the greatest Expected Weight of Evidence for target proficiency
    - Stop when target proficiency learned with specific accuracy (Stopping rule not implemented for field trial)

Weight of Evidence

- Good (1985)
- H is binary hypothesis, e.g., *Proficiency* > Medium
- E is evidence for hypothesis
- Weight of Evidence (WOE) is

$$W(H:E) = \log \frac{P(E|H)}{P(E|\overline{H})} = \log \frac{P(H|E)}{P(\overline{H}|E)} - \log \frac{P(H)}{P(\overline{H})}$$

Conditional Weight of Evidence

- Can define Conditional Weight of Evidence

$$W(H : E_2|E_1) = \log \frac{P(E_2|H, E_1)}{P(E_2|\overline{H}, E_2)}$$

- Nice Additive properties

$$W(H : E_1, E_2) = W(W : E_1) + W(H : E_2|E_1)$$

- Order sensitive
- WOE Balance Sheet (Madigan, Mosurski & Almond, 1997)

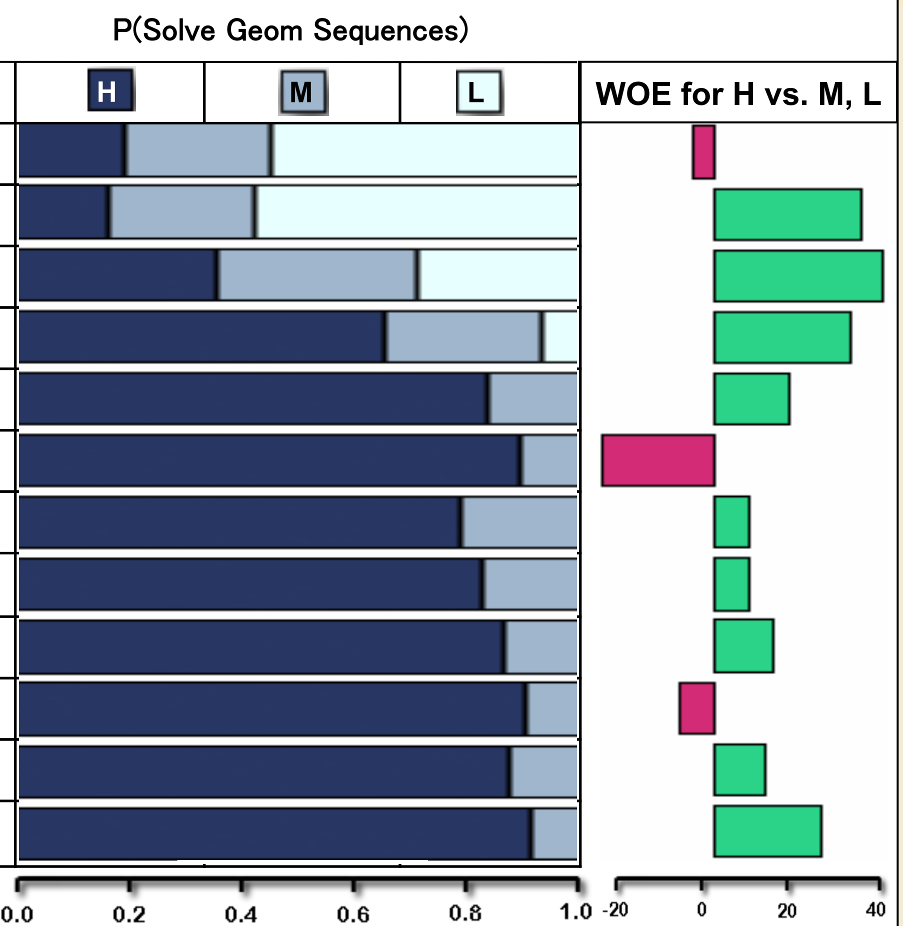
Evidence Balance Sheet

63 tasks total

1 Easy
2 Medium
3 Hard
a Item type
b Isomorph

Task	Acc
SolveGeometricProblems2a	0
SolveGeometricProblems3a	1
SolveGeometricProblems3b	1
SolveGeometricProblems2b	1
VisualExtendTable2a	1
SolveGeometricProblems1a	0
SolveGeometricProblems1b	1
VisualExtendVerbalRule2a	1
ModelExtendTableGeometric3a	1
ExamplesGeometric2a	0
VisualExplicitVerbalRule3a	1
VerbalRuleModelGeometric3a	1

⋮



Expected Weight of Evidence

When choosing next "test" (task/item) look at expected value of WOE where expectation is taken wrt $P(E|H)$.

$$EW(H:E) = \sum_{k=1}^{K} W(H:e_k)P(e_k|H)$$

where $\{e_k : k = 1, ..., K\}$ represent the possible results.

Related Measures

- Value of Information
- Fisher Information
- Mutual Information
  - Extends to non-binary hypothesis nodes
  - Built into Netica
- Kullback-Liebler distance between joint distribution and independence
- Entropy

ACED Evaluation

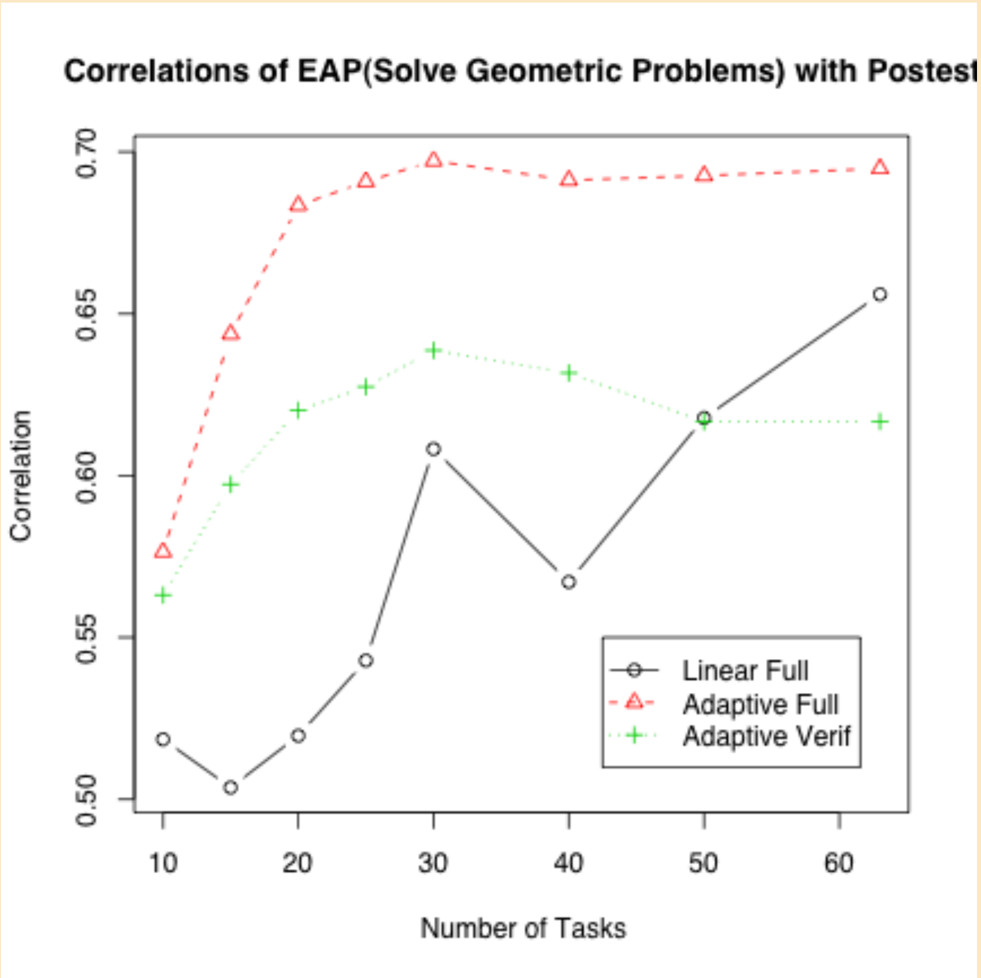
- Middle School Students
- Did not normally study geometric series
- Four conditions:
  - Elaborated Feedback/Adaptive (E/A; n=71)
  - Simple Feedback/Adaptive (S/A; n=75)
  - Elaborated Feedback/Linear (E/L; n=67)
  - Control (no instruction; n=55)
- Students given all 61 geometric items
- Also given pretest/posttest (25 items each)

ACED Reliability

Proficiency (EAP)	**Reliability**
Solve Geometric Sequences (SGS)	*0.88*
Find Common Ratio	0.90
Generate Examples	0.92
Extend Sequence	0.86
Model Sequence	0.80
Use Table	0.82
Use Pictures	0.82
Induce Rules	0.78
Number Right	0.88

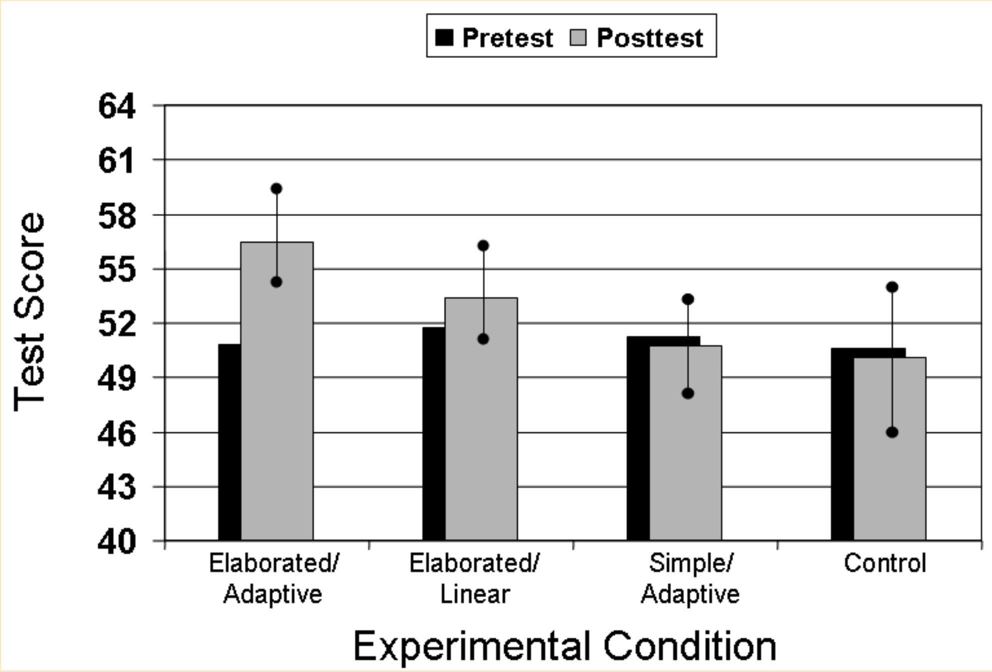
- Calculated with Split Halves (ECD design)
- Correlation of EAP score with posttest is 0.65 (close to reliability of posttest)
- Even with pretest forced into the equation, EAP score accounted for 17% unique variance
- Reliability of modal classifications was worse

Effect of Adaptivity



- For adaptive conditions, correlation with posttest seems to hit upper limit by 20 items
- Standard Error of Correlations is large
- Jump in linear case related to sequence of items

Effect of feedback



- E/A showed significant gains
- Others did not
- Learning and assessment reliability!!!!!

Bigger Picture

- Can use as framework for developing real-time embedded assessment (Stealth Assessment)
- Can score right away using expert judgement for parameters
  - No worse than number right
  - Models can later be calibrated to pretest data (Identifiability issues with some models)
  - Ability to deploy right away helps gather data

Acknowledgements

- ACED development and data collection was sponsored by National Science Foundation Grant No. 0313202.

- Complete data, model specifications and links to papers available at: http://ecd.ralmond.net/ecdwiki/ACED/ACED