UNIVERSITY OF ALBERTA

Augmented Intelligence in Test Development: Using Constraint Programming for Item Generation

Dr. Hollis Lai, Associate Professor
Director of Assessment and Evaluation, Undergraduate Medical Education
Faculty of Medicine and Dentistry
University of Alberta

Dr. Mark Gierl, Professor and Canada Research Chair
Centre for Research in Applied Measurement and Evaluation
University of Alberta

Overview

- **Automatic Item Generation has the ability to produce large numbers of items**
  - Programming or instruction-based approach
  - *Logical constraint and cognitive model-based approach (us!)*
  - Ontology-based or knowledge-base dependent approach (e.g., Mitkov, Gutl or Soltadova)
- **Our approach relies on combinatorial process with constraints to assemble a set of allowable combination of elements to represent a test item**
- **Focus of this talk is limited to how items are generated, for calibrated item generation…**

UNIVERSITY OF ALBERTA

Agenda

1. Where we come from

- Item Generation 101

2. How we do it

- Evolution of how we generate

- Augmented Intelligence: Constraint Programming

3. Applications

- English language items

4. Where we are going

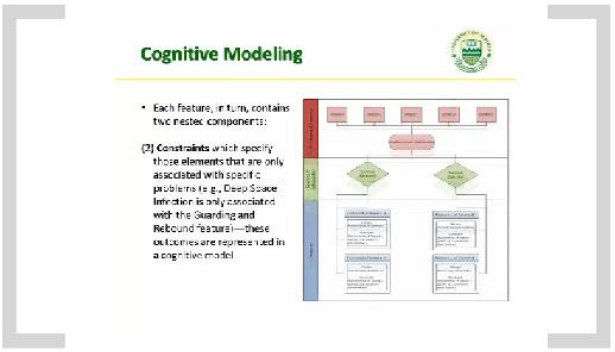
UNIVERSITY OF ALBERTA

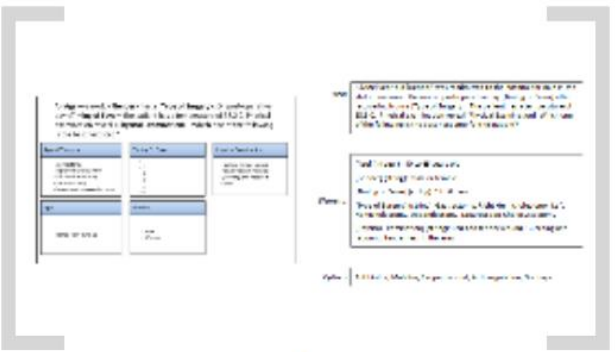
Item Generation 101 – Rationale

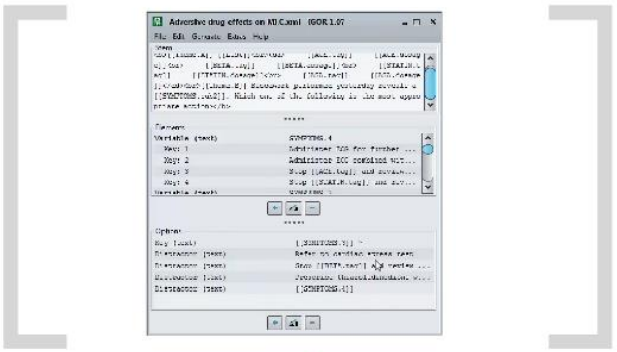
- **The need for items**
  - Computer-based testing, adaptive testing
  - Continuous administration
  - Formative assessment
  - Competency-based assessment

- **Scalability problem**
  - SMEs producing items one at a time
  - Iterative review process (time consuming)
  - Quality and guidelines eliminate items before reaching field test
- **Need for items are increasing drastically while production is scaled by the number of SMEs**

Our Approach







Step 1
Cognitive Modeling

Step 2
Item Modeling

Step 3
Item Generation

An example in surgery

A 54-year-old woman has a laparoscopic cholecystectomy. On post-operative day 3 she has a temperature of 38.5c. Physical examination reveal a red and tender wound and calf tenderness. Which one of the following is the best next step?

a. Mobilize

b. Antibiotics

c. Anti coagulation

d. Reopen the wound

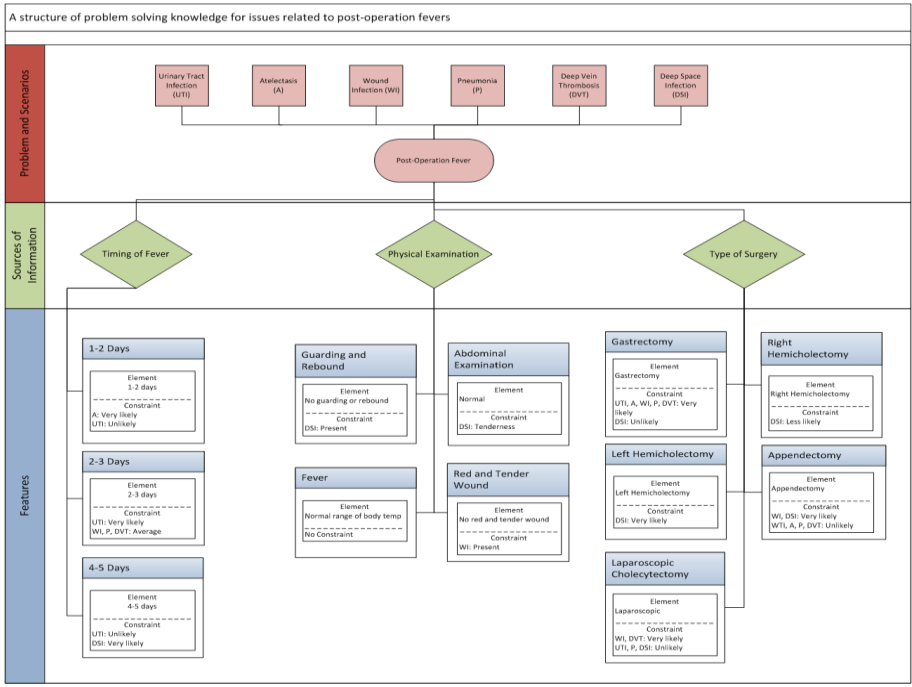
Cognitive Modeling

The model includes three key outcomes:

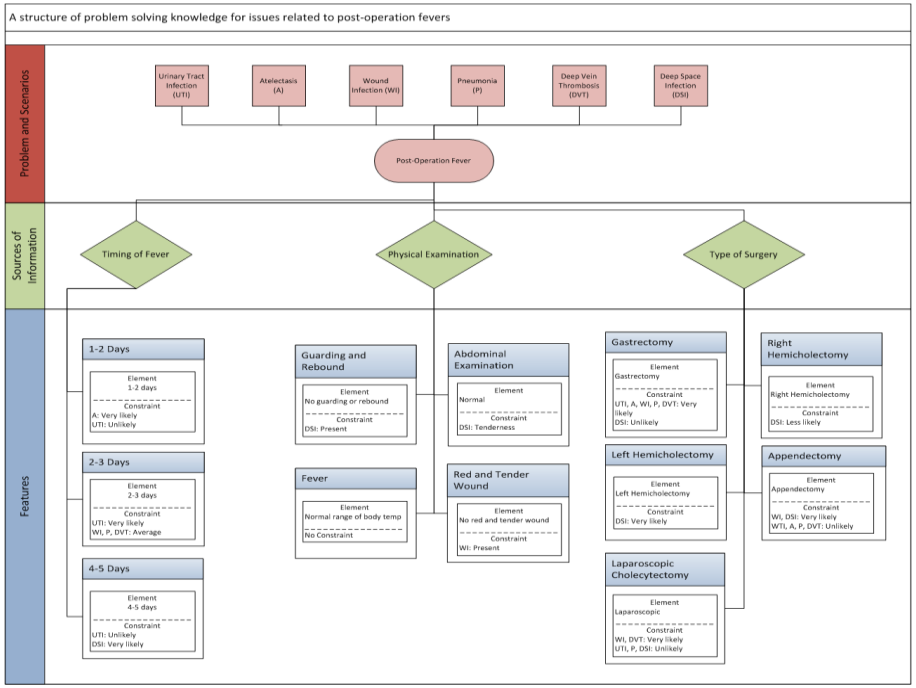
1. Identify **THE PROBLEM** (i.e., Post-Operative Fever);

2. Specify **SOURCES OF INFORMATION** required to diagnose the problem (i.e., Type of Surgery); and

3. Describe **KEY FEATURES** within each information source (e.g., Fever) needed to create different instances of the problem



Item Modeling



A 54-year-old woman has a laparoscopic cholecystectomy. On post-operative day 3 she has a temperature of 38.5c. Physical examination reveal a red and tender wound and calf tenderness. Which one of the following is the best next step?

A 54-year-old woman has a <TYPE OF SURGERY>. On post-operative day <TIMING OF FEVER> the patient has a temperature of 38.5c. Physical examination reveal <PHYSICAL EXAMINATION>. Which one of the following is the best next step?

<TYPE OF SURGERY> Gastrectomy; Right Hemicolectomy; Left Hemicolectomy; Appendectomy; Laparoscopic Cholecystectomy

<TIMING OF FEVER> 1 to 6 days

<PHYSICAL EXAMINATION> Red and Tender Wound; Guarding and Rebound; Abdominal Tenderness; Calf Tenderness

UNIVERSITY OF ALBERTA

Item Generation

- After the item model is specified, we combine this information systematically to produce new items
- To accomplish this combinatoric task, we use a software for item generation called IGOR (Item GeneratOR)
- IGOR was programmed using Sun Microsystems JAVA

igor

1. A 54-year-old woman has a gastrectomy. On post operative day 3 he has a temperature of 38.5 C. Which one of the following is the most likely diagnosis?

   a. Urinary tract infection
   b. Actelectasis
   c. Wound infection
   d. Pneumonia

2. A 54-year-old woman has a gastrectomy. On post operative day 3 he has a temperature of 38.5 C. Which one of the following is the most likely diagnosis?

   a. Urinary tract infection
   b. Actelectasis
   c. Wound infection
   d. Deep vein thrombosis

3. A 54-year-old woman has a gastrectomy. On post operative day 3 he has a temperature of 38.5 C. Which one of the following is the most likely diagnosis?

   a. Urinary tract infection
   b. Actelectasis
   c. Wound infection
   d. Deep space infection

4. A 54-year-old woman has a gastrectomy. On post operative day 3 he has a temperature of 38.5 C. Which one of the following is the most likely diagnosis?

   a. Urinary tract infection
   b. Actelectasis
   c. Pneumonia
   d. Wound infection

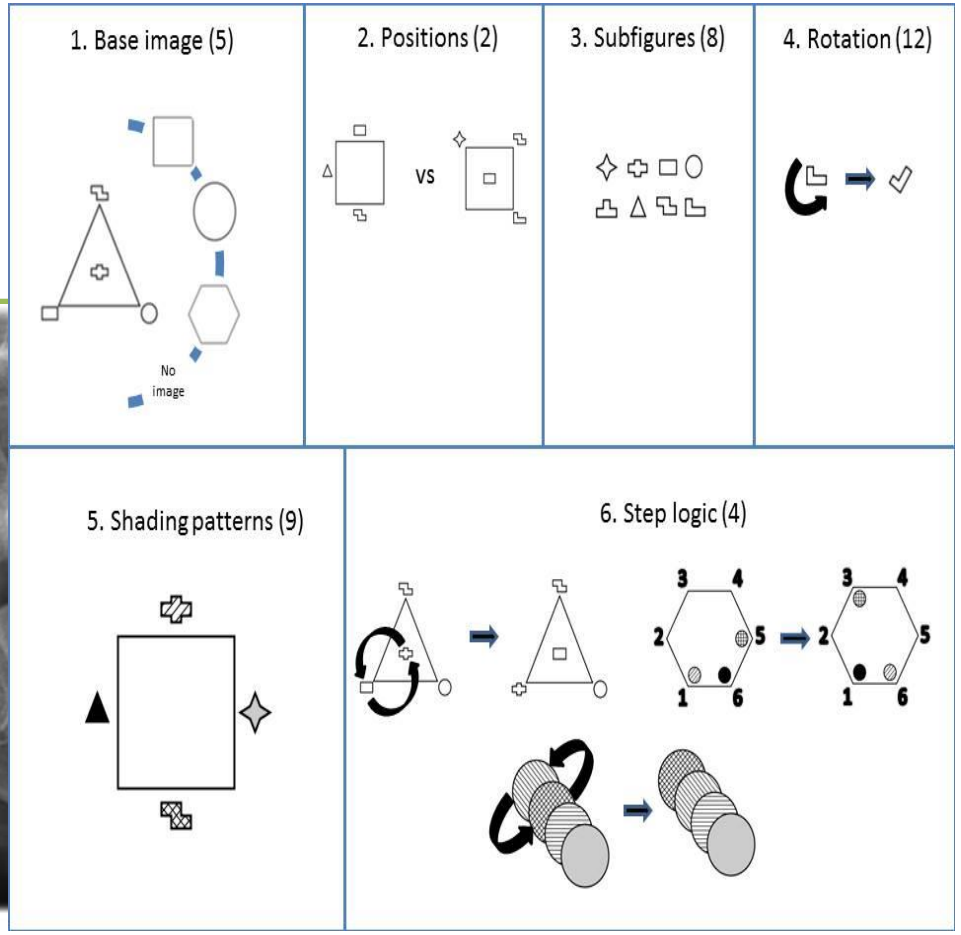
From that process

Mathematics (k-12)

Lai, H., Alves, C. & Gierl, M. (2009)
Lai, H., Gierl, M., Hogan, J. (2015)
Gierl, M. & Lai, H. (2012)

That's great, but what about…?

Topic Area	R
Medicine	Gierl, for mu Medic
Biology	Gierl, for Cc meeti
Non-verbal Reasoning	Gierl, Metho mode Intern Comr Spring
Dentistry	Lai, H (2016 Gene
Science	Gierl, press K-12 (Eds.) skill d



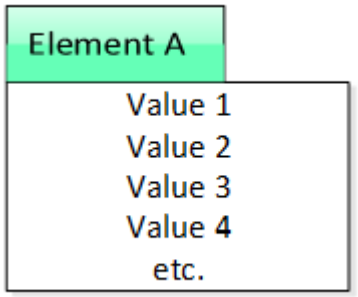
Evolution of how we generate

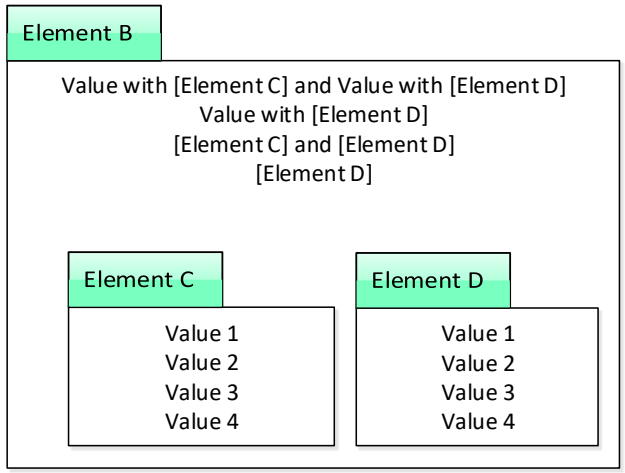
- Application of the three steps has allowed us to generate items
- As we moved into different domains, task of generating items became more complex
- We developed different tools to simplify the task of translating item models for generation

UNIVERSITY OF ALBERTA

Tool 1. N-Layering

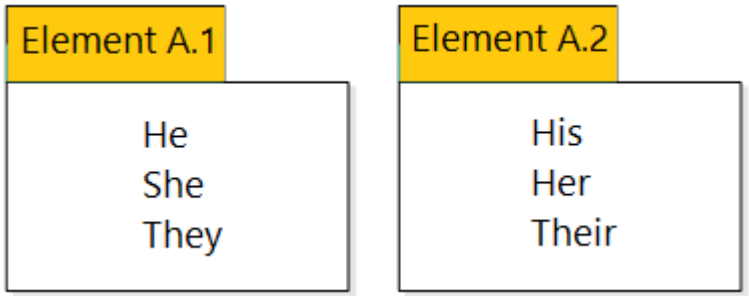
- N-layer permits the manipulations of a non-linear set of generative operations using a large number of elements at a multiple levels
- The concept of n-layer item generation follows development of syntactic structures for item creation (Bormuth 1969)
- Great for expressing item content in a different ways
- Think of this as increasing depth of a model





UNIVERSITY OF ALBERTA

Tool 2. Linked elements

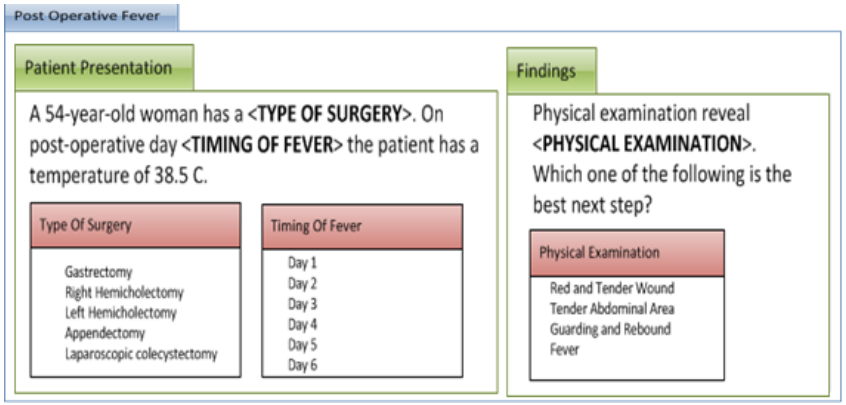


- Linking elements together allows manipulations of one set of elements to be dependent on another
- The concept of linking follows the idea of creating parallel content with a model
- Great for expressing consistent sets of content in items
- Think of this as increasing breadth of a model

UNIVERSITY OF ALBERTA

Using these tools

By stacking various presentations in different layers and creating parallel elements in different languages…



	English	Chinese
Sentence 1	Patient complaints of a mass [[ENG.Pain]] in [[ENG.Location]] which has been a problem since [[ENG.AcuityofOnset]].	**一名患者主訴** [[CH.acuityofonsetCH]]**在** [[CH.locationCH]] **出現的一個** [[CH.painCH]] **包塊。**
Sentence 2	A [[GEN.Gender]] was admitted with pain in [[ENG.Location]] from [[ENG.AcuityofOnset]].	**一名**[[GEN.genderCH]]**子因** [[CH.acuityofonsetCH]][[CH.locationCH]]**出現疼痛而入院。**
Sentence 3	On examination, the mass is [[ENG.PhysicalFindings]] and lab work came back with [[ENG.WBC]].	**經檢查後，那包塊是** [[CH.physicalfindingCH]]**化驗結果顯示**[[CH.WBCch]]**。**

UNIVERSITY OF ALBERTA

Generating items in different languages

21. A 25-year-old woman presented with a mass in the left groin. It occurred a few months ago. On examination, the mass is protruding but with no pain and lab work came back with normal vitals. What is the best next step?

1. ice applied to mass
2. exploratory surgery
3. reduction of mass
4. hernia repair

47. 一名25歲的女患者在左側腹股溝出現一個包塊。徵狀已持續了幾個月。經檢查後，那包塊是突出而不疼痛，化驗 果顯示生命體徵正常。下一部最佳處治是那一個?

1. 在包塊上冷敷
2. 腹腔探查術
3. 手法回復包塊
4. 疝氣修補手術

111. Patient presents with a mass and mild pain in an area near a recent surgery from a few days ago after moving a piano. The patient is a 25-year-old woman. Upon further examination, the patient had normal vitals and the mass is tender and reducible. What is the best next step?

1. ice applied to mass
2. exploratory surgery
3. reduction of mass
4. hernia repair

212. 一名患者的手術的切口附近從數天前，自搬動鋼琴後出現一個有輕微痛感的包塊。患者牲別女，25 歲。經身體檢查後，患者生命體徵正常，而那包塊是軟而可回復的。下一部最佳處治是那一個?

1. 在包塊上冷敷
2. 腹腔探查術
3. 手法回復包塊
4. 疝氣修補手術

UNIVERSITY OF ALBERTA

Tool 3. Expression of Item Constraints

One of our biggest challenge in building item models was to program the constraints such that only plausible items will be generated

As models become more complex, expression of constraints turn into their own scripts

This moves our models back into programming approach for generation, and we need tools to move back into logical constraints

UNIVERSITY OF ALBERTA

Old Constraints

Previously, constraints were expressed Boolean logic statements

A1	Apple		B1	tree
A2	Grape		B2	vine

[[A.value]]==[[B.value]]

This was very time consuming when models become larger and relationships become more complex

```
((([[S5]] == 1) && ([[I1]] > 1)) || ((([[S5]] == 2) && ([[I1]] == 1))
((([[S6]] == 1) && ([[I2]] > 1)) || ((([[S6]] == 2) && ([[I2]] == 1))
((([[S7]] == 1) && ([[I3]] > 1)) || ((([[S7]] == 2) && ([[I3]] == 1))
((([[S8]] == 1) && ([[I4]] > 1)) || ((([[S8]] == 2) && ([[I4]] == 1))
((([[S9]] == 1) && ([[I1]]*[[I3]] > 1)) || ((([[S9]] == 2) && ([[I1]]*[[I3]] == 1))
((([[S10]] == 1) && ([[I2]]*[[I3]] > 1)) || ((([[S10]] == 2) && ([[I2]]*[[I3]] == 1))

[[S11]] == [[S12]]
[[S21]] == [[S22]]
((([[S11]] == [[S21]]) && ([[S41]] == 2) && ([[S42]] == 1)) || ((([[S11]] !=
[[S21]]) && ([[S41]] == 1) && ([[S42]] == 2))
((([[S1]] == [[S2]]) && ([[S3]] == 2)) || ((([[S1]] != [[S2]]) && ([[S3]] == 1))
[[S3]] == 2
([[S32]] == 2) && ([[I2]] != 1)
```

UNIVERSITY OF ALBERTA

New Constraint Programming

We moved into a new form of constraint programming that allows for the expression of relationships without using Boolean logic

For every pair of variables, there is a finite number of relationships of whether a value of one variable can be presented with another

If these relationships are captured in binary, then Boolean logic will not be needed

Models are just a values by values matrix, where at generation, a vector is called to determine where an item should be generated

UNIVERSITY OF ALBERTA

An example of the new constraints

	B1	B2	B3	B4	B5
A1	1	1	0	1	1
A2	1	1	1	0	0
A3	0	0	0	1	0
A4	1	0	0	1	1
A5	1	1	1	1	1

Variable Pair Constraint Matrix

Variable		A 1	A 2	A 3	A 4	A 5	B 1	B 2	B 3	B 4	B 5	C 1	C 2	C 3	C 4
A	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
A	2	1	1	1	1	1	1	1	1	0	0	1	1	1	1
A	3	1	1	1	1	1	0	0	0	1	0	1	1	1	1
A	4	1	1	1	1	1	1	0	0	1	1	1	1	1	1
A	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1
B	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
B	2	1	1	0	0	1	1	1	1	1	1	1	1	1	1
B	3	0	1	0	0	1	1	1	1	1	1	1	1	1	1
B	4	1	0	1	1	1	1	1	1	1	1	1	1	1	1
B	5	1	0	0	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Model Constraint Matrix

Given an item that uses A3,B5 and C3

Pair		Value
A3	B5	0
A3	C3	1
B5	C3	1
Outcome		0

UNIVERSITY OF ALBERTA

New constraint programming benefits

Checkbox interface

No programming required

Lowers complexity of modeling

Much quicker generation time

UNIVERSITY OF ALBERTA

Augmented Intelligence Application

Artificial intelligence is a term for describing the theory and the application of computer systems that perform tasks and solve problems that normally require human intelligence (Russell & Norvig, 2010)

Human-in-the-loop hybrid-augmented intelligence (or *augmented intelligence*, for short) is an AI topic that deals with how computer systems can extend human cognitive abilities thereby improving human task performance (Zheng, Liu, Ren, Ma, Chen, Yu, Xue, Chen, & Wang, 2017)

UNIVERSITY OF ALBERTA

Augmented Intelligence Application

In the past, item modeling was created by the SME while the generation of items was done by programmers (workflow)

With the use of the new constraint programming approach, SME have the potential to naturally express constraints with the model and enable item generation without programming (direct manipulation)

This is a crucial step towards enabling SME with augmented intelligence such that SMEs can improve their ability to generate items

UNIVERSITY OF ALBERTA

Automatic item generation is becoming a symbiotic hybrid process where content specialists are being enhanced by technology to meet the demands of the producing test new test items

The **content specialist** identifies the knowledge and skills required to solve problems and the **test translator** expresses this information across language groups—we associate these activities with the "art" of test development because it requires judgement, expertise, and experience

ART

+

Computer technology is required for the generative task of systematically combining large amounts of information in each item model—we associate this with the "science" of modern computing

SCIENCE

=

By merging the outcomes from the content-based creative task with the technology-based generative task, automated processes can be used to promote a **new approach to item development**

AIG

UNIVERSITY OF ALBERTA

Reading Comprehension Item

At Home in a Language

When I was a child growing up in Delhi, India. My parents and I will have spent our summers in Calcutta, India, visiting my grandparents, aunts, uncles, and cousins. We took the train over eight hundred miles from Delhi to Calcutta, which I considered a treat as itself. I loved the dining car, the cozy sleeping berth in our cabin, and the gentle rocking motion of the train that would lull me to sleep at night. As an adult, I prefer to travel by car. When we arrived at the Calcutta station the next morning, we were welcomed announcing train arrivals and departures over the intercom by the sound of the Bengali language.

Back in Delhi, the language most people commonly spoke was Hindi. Though I spoke Hindi fluently, it wasn't my first language. My parents were born in Calcutta, when most people spoke Bengali. They had lived there for many years before they go married and moved to Delhi, where Hindi was widely spoken. Because my parents had grown up speaking Bengali, we spoke Bengali, not Hindi, in our house. It was not surprising, then, that hearing Bengali on the streets of Calcutta made me feel right at home.

Stem:

When I was a child growing up in Delhi, India. My parents and I will have spent our summers in Calcutta, India, visiting my grandparents, aunts, uncles, and cousins.

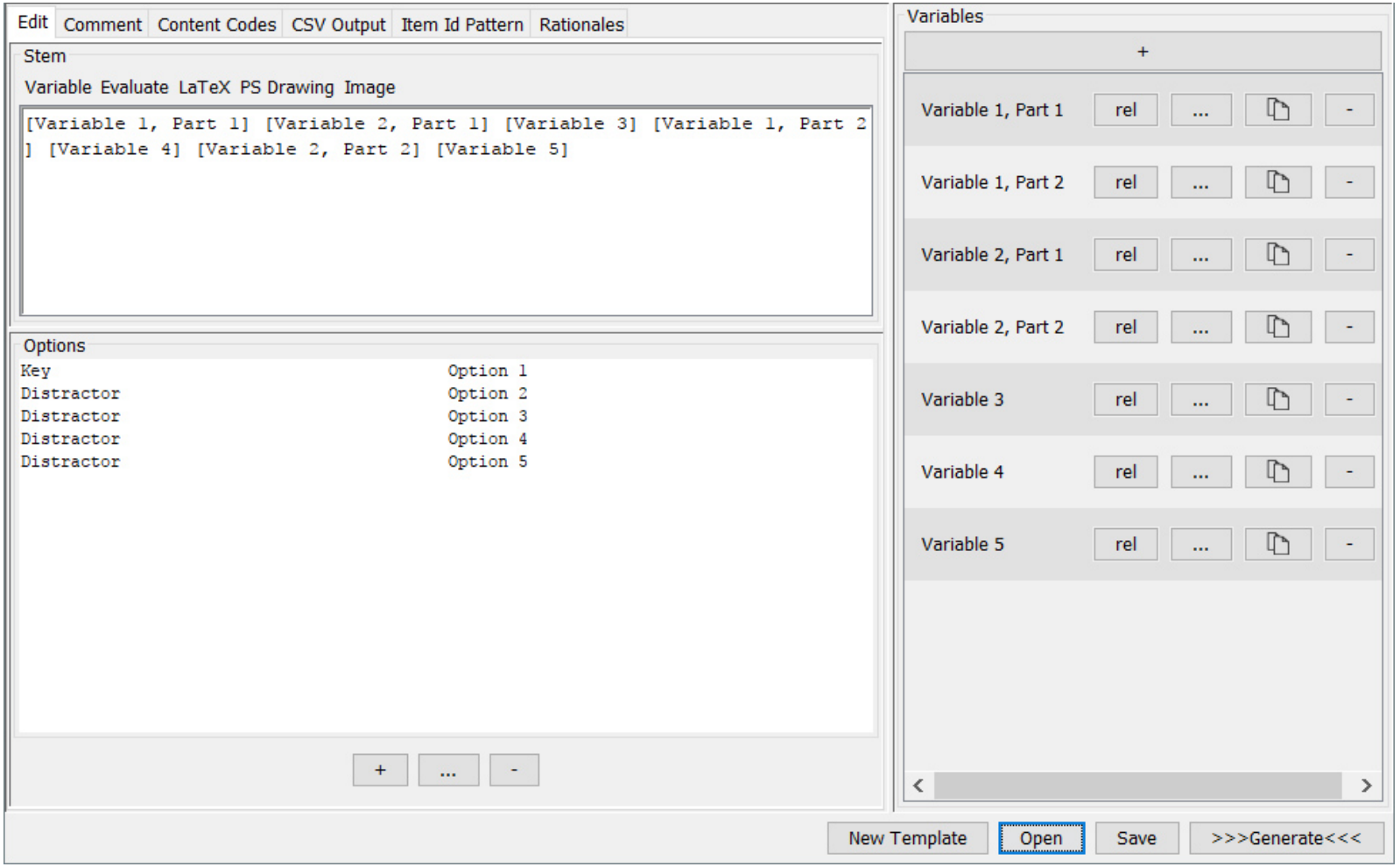
Item Modeling

Item Model Stem:

[Variable 1, Part 1] [Variable 2, Part 1] [Variable 3] [Variable 1, Part 2] [Variable 4] [Variable 2, Part 2] [Variable 5]

Variable 1, Part 1	Variable 2, Part 1	Variable 3		Variable 1, Part 2		Variable 4		Variable 2, Part 2		Variable 5
When I was a child growing up in	Delhi, India,	my parents	and	I would spend	our	summers	in	Calcutta, India,	visiting my	grandparents, aunts, uncles, and cousins.
During the year we lived in	Buffalo, United States,	my family		I spent		weekends		Boston, United States,		grandparents
By the time I leave	Cologne, Germany,	my siblings		I will have spent		holidays		Berlin, Germany,		aunts and uncles
Were I to have stayed in	Tel Aviv, Israel,	my cousins		I would have spent		vacations		Jerusalem, Israel,		cousins
				I spending						
				I will spend						
				I will be spending						
				I have spent						
				NO CHANGE						

Item Generation



From 5832 possible combinations, 83 items generated

Generating English Language Arts Items

The same three step of AIG can be applied to generate items to meet the demands for ELA

New constraint programming enables generation of items in this area

While generation output is smaller in this area, there are lots of opportunities to modify and adapt the modeling approaches to fit the required assessment tasks

Future Directions

- **Expansion the generative capacity in ELA**
- **Enhancing the quality control process of AIG**
  - Developing guidelines that suit SMEs for AIG
  - Moving QC to the model level
- **Further integrating SME within the process of AIG**
  - Becoming more involved in the generation process
  - Integrating AIG into the item development workflow
  - Minimizing the outcomes from scalability

Summary

- **Automatic Item Generation has the ability to produce large numbers of items**
- **Our approach relies on combinatorial process with constraints to assemble a set of allowable combination of elements to represent a test item**
- **With emergence of tools and aids that help SME, AIG is becoming an application of augmented intelligence in test development**
- **Methods for how items will generate items will continue to evolve, AI will continue to close the link between knowledge expression from SME and knowledge prompts for learners.**

Thank you!

Email:

hollis.lai@ualberta.ca

UNIVERSITY OF ALBERTA