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

Advances in Automatic Item Generation with Demonstration

TAO Days—Swiss Conference of Cantonal Ministers of Education
Bern, Switzerland—October 1-2, 2013.
AUTOMATIC ITEM GENERATION

• The benefits of computer-based testing are well documented—CBT permits testing on-demand; CBTs can be administered via the internet eliminating the need for paper; CAT, one popular form of CBT, shortens the exam length resulting in a dramatic reduction in testing time

[THE WORLD IS GOING CBT]

• But CBT also requires large item banks containing thousands of multiple-choice items

• And even if CBT is not used, items for any testing program are becoming exceedingly expensive to create: Rudner (2010) claimed that the cost of developing a single item for a large-scale test ranged from $1500 to $2500
ONE WAY TO CREATE TEST ITEMS...
Another way to address this item development challenge is with automatic item generation (AIG).

Automatic item generation is the process of using item models to generate test items with the aid of computer technology—with this approach, hundreds or even thousands of items can be generated with a single item model.

While the idea of automatic item generation may be viewed as a “dream come true” by many testing agencies, I am here to tell you that the dream is well within our reach because of recent advances in educational measurement theory and practice.
AUTOMATIC ITEM GENERATION

AIG METHODOLOGY
A 54-year-old woman has a laparoscopic cholecystectomy. On post-operative day 3 she has a temperature of 38.5°C. 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
AIG METHODOLOGY
That complex diagram is a cognitive model highlighting the knowledge, skills, and content required to make a medical diagnosis.

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.
A structure of problem solving knowledge for issues related to post-operation fevers

**Problem and Scenarios**
- Urinary Tract Infection (UTI)
- Atelectasis (A)
- Wound Infusion (WI)
- Pneumonia (P)
- Deep Vein Thrombosis (DVT)
- Deep Space Infection (DSI)

**Sources of Information**
- Timing of Fever
- Physical Examination
- Type of Surgery

**Features**
- 1-2 Days
  - Element: 1-2 days
  - Constraint: A: Very Likely, UTI: Unlikely
- 2-3 Days
  - Element: 2-3 days
  - Constraint: UTI: Very Likely, WI, P, DVT: Average
- 4-5 Days
  - Element: 4-5 days
  - Constraint: UTI: Unlikely, DSI: Very Likely

**Guarding and Rebound**
- Element: No guarding or rebound
- Constraint: UTI: Present

**Abdominal Examination**
- Element: Normal
- Constraint: UTI, A, WI, P, DVT: Very Likely, DSI: Unlikely
- DSI: Tenderness

**Gastrectomy**
- Element: Gastrectomy
- Constraint: UTI, A, WI, P, DVT: Very Likely, DSI: Unlikely

**Left Hemicholectomy**
- Element: Left Hemicholectomy
- Constraint: WI, DSI: Very Likely
- DSI: Unlikely

**Appendectomy**
- Element: Appendectomy
- Constraint: WI, DSI: Very Likely
- UTI, P, DSI: Unlikely

**Laparoscopic Cholecystectomy**
- Element: Laparoscopic
- Constraint: UTI, DSI: Unlikely
Next, the item models are created using the cognitive model content, where an item model is like a template, a rendering, or a mould of the assessment task (i.e., it’s a target where we want to place the content for the item).

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 Hemicolecotomy, Left Hemicolecotomy, Appendectomy, Laparoscopic Cholecystectomy

**TIMING OF FEVER:** 1 to 6 days

**PHYSICAL EXAMINATION:** Red and Tender Wound, Guarding and Rebound, Abdominal Tenderness, Calf Tenderness
After the item model is specified, we combine this information systematically to produce new items.

To accomplish this complex combinatoric task, we created software for item generation called IGOR (Item GeneratOR).

IGOR was programmed using Sun Microsystems JAVA.
When we used our three-stage process with 5 different surgery item models, more than 20,000 items were generated.

Item Model 1: Gallstones—288
Item Model 2: Hernias—256
Item Model 3: Aneurism—5,184
Item Model 4: Post Operation Management—7,488
Item Model 5: Post Operation Fever—7,680

TOTAL NUMBER OF GENERATED ITEMS: 20,896
Automatic item generation is the process of using item models to generate test items with the aid of computer technology—it requires a merger between the “art” and “science” of item development.

The test developer identifies the knowledge and skills required to solve problems, organizes this information into a cognitive model, and designs item models—we associate these activities with the “art” of test development because it requires judgement, expertise, and experience.

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.

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.
THANK YOU

Dr. Mark J. Gierl (mark.gierl@ualberta.ca)