FROM SMART TESTING TO
SMART LEARNING

Hua-Hua Chang
University of Illinois at Urbana-Champaign

December 14, 2017
Hong Kong
What is smart learning?
Smart Education - Emergence
Smart Education - Concepts

• “Creating a smart environment for learning that promotes the development of both the learner’s well-roundedness and specialized competency, creating value for the entire society.” (Zhu & Shen, 2013)

• A few aspects (IBM):
  • Technology immersion
  • Individualized learning paths
  • Demand oriented knowledge/skills
  • Globally integrated & accessible systems/resources
  • Leading role of ed in society
Smart Education – Innovations and Attempts

• Personalized learning routes
  • Knewton system
• Blended learning
  • Khan Academy
  • Intelligent Tutoring Systems
  • e-Schoolbags/student portfolios
• Open Educational Resources and Massive Online Open Courseware
  • Coursera
  • EdX
• International expansions and personal device use
Knewton System --- Personalized Learning Paths
“A is for Adaptive-- Personalized learning is poised to transform education. Can it enrich students and investors at the same time?” (Time, June 17, 2013)

Bradley: “Knewton is based on what you can do, not what the class can do”  
Clenna: “It adapts to you, so it starts easy and then gets harder”
MOOCs

• MIT OCW and the OER initiative
• Massive Online Open Courseware (interactive learning):
  • Instruction
  • Resource provision
  • Peer grading
  • Forum discussion
  • Certification
• edX, Udacity, Coursera (>10 mil users, 839 courses, 114 institutes as of 2014)
• Personal device and mobile interfaces
• Self-sufficient
Challenges for Smart Education

• High drop-out for MOOCs; Possible reasons:
  • Audience: adults with scarce time
  • Uniform starting point for all participants
  • Low cost --- low commitment

• Digital divide (New York Times, 2012)
  • Inappropriate usage of digital devices

• Conflict with performance-based personnel selection

• In common: Time scarcity and resource overwhelm

• Question: How do we reliably identify individuals’ needs and filter the resources tailoring individuals’ demands?
Smart Testing

Traditional Testing
- No careful evaluation
- Linear test
- Test/group specific score
- Total score summarizing performance
- Usually based on Classical Test Theory

Smart Testing
- Theoretically sound
- Adaptive test
- Generalizable score
- Provides additional information
- Based on modern test theory (eg. IRT)
Methodology development: Computer adaptive testing (CAT)

A main component in smart testing
Stochastic Approximation

- Stochastic approximation (Robbins & Monro, 1951) provides a systematic way on how the adjustment should be made.
- It has had fundamental impacts on
  - Stochastic control of engineering systems
  - Toxicity studies in drug development
  - Computational statistics
  - Educational testing
  - ...
- It’s backed by mathematical theory.
Stochastic Approximation
Robbins-Monro Process (1951)

Responses: \( x_1, x_2, x_3, \ldots \)
Design points: \( b_1, b_2, b_3, \ldots \)
Constants: \( \delta_1, \delta_2, \delta_3, \ldots \)

\[
b_{n+1} = b_n - \delta_n x_n
\]

\( b_n \rightarrow m \) (a point of interest)
The idea of Robbins-Monro stochastic approximation was extended by F. Lord to CAT.

Key modification of the stochastic approximation:

Parametric modeling (e.g., Logit) of response curves, which is crucial to small sample/initial learning of examinees' response curves.
Example of research problem
Mathematical Foundation of MIC

Under Local Independence (LI),

\[ \hat{\theta}_n \rightarrow \theta_0 \text{ as } n \rightarrow \infty \text{ with } \text{var}(\theta_n) \rightarrow \frac{1}{I(\theta_0)} \]

Thus, the closeness of \( \hat{\theta}_n \) to \( \theta_0 \) is governed by \( I(\theta) \).

In CAT, LI may not hold (Mislevy & Chang, 2000).

How can we establish a mathematical foundation for information-based CAT estimation?
Interplay with Educational Testing

\(X_1, \ldots, X_j, X_{j+1}, \ldots\)

• \(\hat{\theta}_j, \hat{\theta}_{j+1}, \ldots\)

• For 2PL,

\[
\hat{\theta} \text{ is solved from } \sum_{i=1}^{j} a_i(X_j - P_j(\theta)) = 0
\]

\(\mathcal{F}_j \subset \mathcal{F}_{j+1}\)?

Yes! \(\{X_1, X_2, \ldots, X_j\} \subseteq \{X_1, X_2, \ldots, X_{j+1}\}\)

Results:

\(\{X_k - P_k(\theta)\}\) is a martingale difference sequence
Chang & Ying (2009):

\[ (3.4) \quad \sqrt{\sum_{i=1}^{n} a_i^2 (\hat{\theta}_n - \theta)} \rightarrow_{\mathcal{L}} N(0,1). \]

The normalizing factor \( \sqrt{\sum_{i=1}^{n} a_i^2} \) in (3.4) may be replaced by \( \sqrt{I^{(n)}(\hat{\theta}_n)} \) or \( \sqrt{I^{(n)}(\theta)} \), where

\[ (3.5) \quad I^{(n)}(\theta) = \sum_{i=1}^{n} a_i^2 \frac{e^{a_i(\theta - b_i)}}{[1 + e^{a_i(\theta - b_i)}]^2} \]

is the observed Fisher information.
From Theory to Application

• Issues with Max-Inf. methods:
  • High-a items are always preferred by CAT
  • Item pool usage
  • Test security
  • Early estimation of theta may not be stable
    • Chang & Ying (2009) divergence example

• Some designing components should be added to the Max-Inf. Methods!
Thanks to Browser/Server (B/S) Architecture

The cutting-edge B/S architecture allows schools to implement CAT with little to no additional cost using their current computer labs and networks.
HSK web-based adaptive testing system

HSK --- Chinese Language Proficiency Test, the world second largest language proficiency examination
A Large Scale Smart Testing Project

- Developing Cutting-edge technology for HSK (Chinese Proficiency Level Examination, the world’s second largest language proficiency exam)
  - Computerized Adaptive Testing
  - Cognitive Diagnosis
  - CD-CAT

- Supported by Chinese Testing International of Beijing, we currently are leading a group of graduate students in developing a CD-CAT version of the Level 4 HSK.
Why adaptive testing?

(1) Paper-pencil test

• **Reason 1:** Select the most suitable items for individual, reduce test length and burden

![Diagram showing the selection of items for different ability levels](image)
Why adaptive testing?

- **Reason 1**: Select the most suitable items for individual, reduce test length and burden

![Adaptive item pool diagram]

- Difficult items
- Medium difficulty items
- Easy items

- High ability
- Medium ability
- Low ability
Why adaptive testing?

- Reason 2: Large item pool, low test overlap, reduce risk for cheating and security breach

<table>
<thead>
<tr>
<th>Paper-pencil Test</th>
<th>Adaptive Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% test overlap rate</td>
<td>Questions selected based on individual’s previous answers, highly random</td>
</tr>
<tr>
<td>One compromised test form threatens the entire test</td>
<td>Large item pool, resilient to small number of leaked item</td>
</tr>
<tr>
<td>Students receive same test questions, easier to share answers during test</td>
<td>Students receive different questions, hard to cheat</td>
</tr>
</tbody>
</table>
HSK4 web-based Adaptive testing demo --- Student interface

• 10-item demo test, begins after examinee login
HSK4 WEB-BASED ADAPTIVE TESTING DEMO --- STUDENT INTERFACE

- Automatically generate student report when test terminates

Note: The diagnostic report is still under development, and due to time limit, a 10-item test cannot provide accurate estimate on fine-grained skills. Hence we use a sample report here.
From adaptive testing to Personalized learning
CAT helps learning

• The idea of personalized learning is not new. But before technology is ready it is impossible to provide 1-to-1 teaching on a large scale.

• CAT can help!
  • Selecting items sequentially helps students better understand the concepts being taught
  • CAT provides more flexibility

• Examples
  1. Some schools in China are using CAT to help classroom teaching
  2. At UIUC, CD-CAT has been using to help low-performing students in an undergraduate physics course.
Cognitive Diagnostic CAT
What is reported to examinees?

Traditional Testing:

Cognitive Diagnosis:

\[ \theta \]

\[ \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_K] \]

A single score

A set of scores:
One for each attribute.

(\(K\) is the total # of attributes.)

Combine latent trait models with latent class models
Why is this beneficial?

Feedback from an exam can be more individualized to a student’s specific strengths and weaknesses.

\[
\hat{\alpha} = [0000111]
\]

\[
\hat{\alpha} = [0101100]
\]
The Item-Attribute Relationship

Which items measure which attributes is represented by the Q-matrix:

\[
\begin{bmatrix}
i_1 & i_2 & i_3 & i_4 \\
A1 & 0 & 1 & 0 & 1 \\
A2 & 1 & 0 & 0 & 1 \\
A3 & 1 & 0 & 1 & 0
\end{bmatrix}
\]
IRT Models

• Many models were proposed
  • DINA model
  • Fusion model

• Methodology development: item selection based on latent class alpha

\[ P(X_{ij} = 1 \mid \alpha_i) \]
CD-CAT supports *all-the-time and everywhere* learning

- The assessment components in most on-line learning systems are not smart enough
  - “adaptive” or “tailor-made”?
  - Termination rule?
  - Diagnostic report?
  - How to detect learning?

- Learners should be further encouraged and inspired

- Also, shall we get rid of teachers in classrooms?
In December 2011, 30,000 Grade 5 Students in Dalian China were taking a cognitive diagnostic CAT for their English proficiency assessment.
Utilizing CAT in Classroom Teaching, Students are learning “Area of a Circle”

“圆的面积” 课例展示（北京市海淀区西颐小学六年级二班）

图片说明：1. 集体学习系统中“圆的面积”的视频内容；
Students really enjoy the new mode of testing, which makes learning more enjoyable comparing with regular teaching and P&P testing.
Validity study

- Compare CD-CAT output with
  - (1) students’ English achievement exam
  - (2) teachers’ evaluation

| TABLE 11.5 The Consistency of Performance Levels With the Number of Mastered Attributes |
|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| Academic Performance Level | The Number of Mastered Attributes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| Excellent                      | 0                              | 0                              | 1                              | 1                              | 1                              | 3                              | 4                              | 6                              | 23                              | 39                              |
| Good                           | 0                              | 0                              | 1                              | 2                              | 8                              | 5                              | 7                              | 7                              | 5                               | 33                              |
| Basic                          | 1                              | 1                              | 3                              | 5                              | 3                              | 1                              | 0                              | 0                              | 1                               | 15                              |
| Below basic                    | 0                              | 1                              | 2                              | 0                              | 0                              | 0                              | 0                              | 0                              | 0                               | 3                               |
| Total                          | 1                              | 2                              | 7                              | 8                              | 12                             | 9                              | 11                             | 13                             | 27                              | 90                              |
Most Students said the CD-CAT is helpful!

Teachers (郑州金水实验区)：Assigning different items to each student, CAT encourages critical thinking, and makes students more independent in problem solving, and offers remedy according to their individual needs, which makes learning more interesting.

Is CD-CAT helpful to your learning?

- greatly
- yes
- no
How many times/week you use CD-CAT without teacher’s assignment

- never, 47, 26%
- once, 40, 22%
- 2 times, 39, 21%
- 3 and above, 58, 31%

How many minutes each time

- < 20 minutes, 49, 24%
- 20-40 minutes, 112, 55%
- 40-60 minutes, 27, 13%
- > 60 minutes, 15, 8%
Help Teachers Know Their Students Better. According to the diagnostic report, remedial planning is on the way. The in-class CAT provides more information to teachers, which facilitates research and career development.
Example: Help classroom learning

The purpose is not to get rid of teachers. Teachers can teach more effectively
Another Example in China (Dalian, China): Adaptive Testing Administered By Paper/pencil

How to make P&P Test Adaptive?
Example CAT in STEM Fields

• High Drop-out Rates in STEM Field
  • Poor exam performance cause many students to drop out of science & engineering curricula, lowering retention rates in STEM disciplines.
  • Lower performing (yet well-prepared) students are poor at predicting their own performance before, (and even after) taking exams, and hence are at risk of failing and eventually dropping out of STEM majors.

• We use CD-CAT to help low performing students in a physics course!
CAT in STEM Fields

• Research Design in the UIUC Physics Course

Hourly Exam 1
Mid-term P&P

Hourly Exam 2
Mid-term P&P

Hourly Exam 3
Mid-term P&P

Hourly Exam 4
Final P&P

Identify students scoring below 70%

Three CD-CATs

Three CD-CATs

Pass

Fail

✓ Intrv 1: CD-CAT
✓ Intrv 2: CD-CAT + Worked Examples
✓ Intrv 3: CD-CAT + Interactive problems
✓ Intrv 4: CD-CAT + Human tutoring
✓ Control group
Example CD-CAT in STEM Fields

- Project Stages

1. Collect & Administer high-stakes in-class tests
2. Develop a web-based CD-CAT platform
3. Build item pools by analyzing legacy items
4. Recruit students and administer CD-CAT
5. Provide remedial interventions based on CD-CAT results
6. Collect & Analyze data
Example CD-CAT in STEM Fields

- Example Data Coding

<table>
<thead>
<tr>
<th>Item</th>
<th>Solving problems using trigonometry</th>
<th>Solving conceptual problems</th>
<th>Solving problems using formulas</th>
<th>Selecting the appropriate variables</th>
<th>Differentiating between energy and work-energy</th>
<th>Differentiating between events</th>
<th>Solving problems involving a change in reference frames</th>
<th>Solving problems involving more than one subconcept</th>
<th>Content area</th>
<th>Answer key</th>
<th>Number of choices</th>
<th>Paired</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example CD-CAT in STEM Fields

• Snapshot of CD-CAT Web-Delivery

Login page via university course website
Example CD-CAT in STEM Fields

• Snapshot of CD-CAT Web-Delivery

A block of mass $m=2.8 \text{ kg}$ is initially at rest at the bottom of a frictionless ramp. A horizontal force $F=10\text{ N}$ pushes the block to the top of the ramp. The height of the ramp $H=1 \text{ m}$, and the ramp’s angle $\theta=20^\circ$.

- $27.5\text{ m/s}^2$
- $40.0\text{ m/s}^2$
- $27.6\text{ m/s}^2$
- $41.8\text{ m/s}^2$
- $65.7\text{ m/s}^2$

What is the $y$ component of the acceleration of the cart at the instant shown?

- $4.8 \text{ m/s}^2$
- $5.1 \text{ m/s}^2$
- $8 \text{ m/s}^2$
- $9.3 \text{ m/s}^2$
- $9.8 \text{ m/s}^2$

Student’s mode of taking CD-CAT
Example CD-CAT in STEM Fields

- Snapshot of CD-CAT Web-Delivery

Administrator’s mode of maintaining CD-CAT
2014-2017, Students graduated all landed great jobs!

- Dr. Yi Zheng, Assistant Professor, Arizona State University
- Dr. Shiyu Wang, Assistant Professor, University of Georgia
- Dr. Chanjin Zheng, Distinguished Professor, Jiangxi Normal University
- Dr. Rui Guo, Data Scientist, Kahuna Inc., CA
- Dr. Cong Chen, Data Analyst, East Side Union High School District, CA
- Dr. Poh Hua Tay, Deputy Director of Assessment, Ministry of Education, Singapore
- Dr. Shuai Wang, Research Scientist, SRI International, CA
- Dr. Hyeon-Ah Kang, Assistant Professor, University of Texas at Austin
- Dr. Sam Ye, Assistant Professor, University of Missouri – Kansas City
- Dr. Justin Kern, Assistant Professor, University of California, Merced
- Dr. Edison Choe, Research Manager, GMAC, Reston, VA
From smart testing to smart learning: how testing technology can assist the new generation of education

Susu Zhang*
University of Illinois at Urbana-Champaign,
603 E Daniel St., Champaign, IL 61820, USA
Email: szhanc105@illinois.edu

Hua-Hua Chang*
University of Illinois at Urbana-Champaign,
603 E Daniel St., Champaign, IL 61820, USA
and
East China Normal University,
3663 North Zhongshan Road, Shanghai, 200062, China
Email: hhchang@illinois.edu
*Corresponding authors

Abstract: This paper provides a brief overview of the theories and research in modern measurement and test theory, also known as smart testing, and further discusses how smart testing relates to the concepts and practices in smart learning. An introduction to smart learning and some of the challenges it is facing will be provided, followed by an introductory survey to a selected few topics in psychometrics, such as item response theory, computerised adaptive testing, large-scale assessment, cognitive diagnosis, and linking, etc. A couple of models regarding the implications of smart testing theories and techniques in smart education will then be proposed, together with the descriptions of some of the ongoing projects combining the two, as well as a discussion on potential future research directions.

Keywords: computerised adaptive testing; CAT; cognitive diagnosis; personalising online learning; mobile learning; smart testing; smart learning; massive online open courses; MOOCs.


Biographical notes: Susu Zhang is a PhD student in Quantitative Psychology at the University of Illinois at Urbana-Champaign. She received her BA in Psychology from Bryn Mawr College and her MA in Mathematics from Haverford College, Pennsylvania, USA. Her primary research interests are item response theory, computerised adaptive testing and cognitive diagnosis.

Hua-Hua Chang is a Professor of Educational Psychology, Psychology and Education at the University of Illinois at Urbana-Champaign. He is the Editor-in-Chief of Applied Psychological Measurement, former President of the Psychometric Society (2011–2013) and a Fellow of the American Psychological Association.