An Intelligent CAT That Can Deal With Disengaged Test Taking

Steven L. Wise, NWEA

Presentation at the 17th Annual Maryland Assessment Conference: Application of Artificial Intelligence (AI) to Assessment
Definition of a CAT

- **Computer-adaptive tests** are designed to adjust their level of difficulty—based on the responses provided—to match the knowledge and ability of a test taker. [The Glossary of Education Reform]
Artificial Intelligence (AI) and CAT

- Computerized Adaptive Testing (CAT)
  - Cited as the first application of AI in testing
  - Combines IRT with the branching capability offered by computer-based testing
  - The AI in a CAT is focused on item difficulty
    - Matching item difficulty to test taker ability
    - Efficient testing; more motivating (possibly)
Forty Years of CAT

CATs began to appear in the 1970’s, and became more prevalent as desktop computers became more common.

But since their beginning, CATs have shown limited evolution in AI beyond the original emphasis on efficiency.

Examples of tweaks to the basic CAT idea:
  - Exposure control
  - Item enemies
  - Content balancing (alignment)
  - Information balance through test length
Why So Little Evolution?

- Largely due to the fact that most computer-based tests (CBTs) emerged from operational paper-and-pencil programs.
- Led to concerns about mode effects, which constrained what CBTs could do.
- However, CBTs are beginning to become less tethered to paper-and-pencil versions.
- This invites the question: What can CBTs (and, by implication, CATs) become?
Reconceptualizing Adaptivity More Broadly

- A CAT could adapt in other ways than simply adjusting item difficulty.

- **New Definition:** A CAT is a type of computer-based test that can adapt, during a test event, to test taker behavior in a way that can improve test efficiency and the validity of the score(s) produced.

- For example, the impact of construct-irrelevant factors could be reduced.

- Construct-irrelevant factor of interest: test-taking engagement
The Problem of Disengaged Test-Taking

+ Test takers sometimes become disengaged during a test event, even though our measurement models assume they don’t.
+ That is, the test taker does not try to apply his/her knowledge, skills, or abilities to answer at least some of the items.
+ Disengagement can seriously distort test scores and threaten validity.
+ Item response time helps us identify this behavior (rapid guessing).
**Effort Monitoring**

- A CBT can monitor test taker engagement by detecting, in real time, rapid-guessing behavior.
- If this behavior is detected, some type of intervention could be implemented.
- The goal of the intervention is to promote score validity, by preempting additional disengagement.
The Effort-Monitoring CBT (2006)

+ If 3 consecutive rapid guesses occurred, the following message would pop up on the test taker’s computer screen:
  - *Your responses to this test indicate that you are not giving your best effort.*
  - *It is very important that you try to do your best on the tests you take on Assessment Day. These assessment data are used by <the university> to better understand what students learn at <the university>, and what improvements need to be made. In addition, <the university’s> assessment data are reported to the state as evidence of what <the university’s> students know and can do.*

+ If disengagement re-occurred, a more strongly worded 2nd message was given.
Experimental Studies

- Test takers were randomly assigned to CBT versions that either did or did not give message to test takers exhibiting disengagement.
- University general education assessment in scientific reasoning
- Messages had a positive effect on both engagement and test performance.
Experimental Results

- **Study 1** (Wise, Bhola, & Yang, 2006)
  - Findings for those deserving first message:
    - Engagement (RTE) increased ($p < .001$; Standardized ES = .78)
    - Performance increased ($p = .09$; ES = .32)
    - Higher correlations of test performance with SAT scores.

- **Study 2** (Kong, Wise, Harmes, & Yang, 2006)
  - Findings for those deserving first message:
    - Engagement (RTE) increased ($p < .001$; ES = 1.37)
    - Performance increased ($p < .001$; ES = .61)
    - Higher correlations of test performance with SAT scores.
Disengagement in CATs

- CATs are as vulnerable to disengaged test taking as other CBTs.
- They do, however, have an additional problem.
- Disengagement can confuse the item selection algorithm.
- If a test taker disengages for a sizable set of items and then re-engages, upon re-engagement the item difficulty can be severely mistargeted.
- Switching between engagement and disengagement is not uncommon.
Typical Example From the MAP®Growth™ Assessment
Disengagement, Then Re-engagement

![Graph showing item positions and RIT values](image)
A New, Smarter CAT (MAP Growth)

MAP Growth can now, in real time, adapt in two innovative ways:

- Preempt rapid guessing through messaging. Notifications will be sent to proctors rather than test takers.
- Ignore rapid guesses when calculating provisional achievement estimates (used to select items).
## TEST STUDENTS

**Testing Session Name:** 105 Session Password: 7746

<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Status</th>
<th>Approximate Question #</th>
<th>Test Assignment</th>
<th>Accommodations</th>
<th>Disengaged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams</td>
<td>Debbie</td>
<td>Testing</td>
<td>17</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Asmoy</td>
<td>Cindy</td>
<td>Testing</td>
<td>15</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Barclay</td>
<td>Stan</td>
<td>Testing</td>
<td>8</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Cassidy</td>
<td>Ten</td>
<td>Testing</td>
<td>11</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Conner</td>
<td>Jane</td>
<td>Testing</td>
<td>7</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Douglas</td>
<td>Jim</td>
<td>Testing</td>
<td>8</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Horneis</td>
<td>Tenis</td>
<td>Testing</td>
<td>12</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Hannington</td>
<td>Pat</td>
<td>Testing</td>
<td>9</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Hernandez</td>
<td>Ten</td>
<td>Testing</td>
<td>7</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Kobol</td>
<td>Samantha</td>
<td>Testing</td>
<td>9</td>
<td>Growth Math 2-6 MS 2007</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

- **Total Students:** 15
- **Testing:** 9
- **Proctor Action Needed:**
  - To Be Confirmed: 0
  - Paused: 0
- **Confirms New:**
- **End Testing Session:**

This page updates every 20 seconds. Click refresh Status to see current information.

Apply a change to multiple students by choosing from the Select Status dropdown, then selecting the change to apply from the Select Action options.
## TEST STUDENTS

**Testing Session Name:** TOS Session Password: 7746

**Proctor Action Needed:** To Be Confirmed

<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Status</th>
<th>Approximate Question #</th>
<th>Test Assignment</th>
<th>Accommodations</th>
<th>Disengaged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams</td>
<td>Debbie</td>
<td>Testing</td>
<td>17</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Asimov</td>
<td>Cindy</td>
<td>Testing</td>
<td>15</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Barkey</td>
<td>Stan</td>
<td>Testing</td>
<td>6</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Castrity</td>
<td>Tim</td>
<td>Testing</td>
<td>11</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Connor</td>
<td>Jane</td>
<td>Testing</td>
<td>7</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Douglass</td>
<td>Jim</td>
<td>Testing</td>
<td>8</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Heinlein</td>
<td>Terie</td>
<td>Testing</td>
<td>12</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Harrington</td>
<td>Pat</td>
<td>Testing</td>
<td>9</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Hernandez</td>
<td>Tom</td>
<td>Testing</td>
<td>7</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Kobol</td>
<td>Samantha</td>
<td>Testing</td>
<td>9</td>
<td>Growth: Math 2-5 MS 2007</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

Apply a change to multiple students by choosing from the Select Status dropdown, then selecting the change to apply from the Select Action options.
## TEST STUDENTS

**Testing Session Name:** Test Session Password: 7746  
**Proctor Action Needed:** To Be Confirmed: 0  
**Paused:** 0

**Total Students:** 15  
**Testing:** 9

### Students Disengaged

- **Cassidy, Tim**  
  - 1 min ago

### Table: Test Assignment

<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Status</th>
<th>Approximate Question #</th>
<th>Test Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams</td>
<td>Debbie</td>
<td>Testing</td>
<td>17</td>
<td>Growth: Math 2-5 MS 2007</td>
</tr>
<tr>
<td>Asimov</td>
<td>Cindy</td>
<td>Testing</td>
<td>18</td>
<td>Growth: Math 2-5 MS 2007</td>
</tr>
<tr>
<td>Cassidy</td>
<td>Tim</td>
<td>Testing</td>
<td>11</td>
<td>Growth: Math 2-5 MS 2007</td>
</tr>
<tr>
<td>Connor</td>
<td>Jane</td>
<td>Testing</td>
<td>7</td>
<td>Growth: Math 2-5 MS 2007 No</td>
</tr>
<tr>
<td>Douglas</td>
<td>Jim</td>
<td>Testing</td>
<td>8</td>
<td>Growth: Math 2-5 MS 2007 No</td>
</tr>
<tr>
<td>Heileen</td>
<td>Terrie</td>
<td>Testing</td>
<td>12</td>
<td>Growth: Math 2-5 MS 2007 No</td>
</tr>
<tr>
<td>Harrington</td>
<td>Pat</td>
<td>Testing</td>
<td>9</td>
<td>Growth: Math 2-5 MS 2007 No</td>
</tr>
<tr>
<td>Hernandez</td>
<td>Tom</td>
<td>Testing</td>
<td>7</td>
<td>Growth: Math 2-5 MS 2007 No</td>
</tr>
<tr>
<td>Kobol</td>
<td>Samantha</td>
<td>Testing</td>
<td>9</td>
<td>Growth: Math 2-5 MS 2007 No</td>
</tr>
</tbody>
</table>

---

**Student Engagement:**
- The 1 student below is skipping through test questions.  
- Encourage the student to make their best effort.
- Refreshes every 20 seconds.

[腳本。]
What More Can be Done in the Future?

Other potential indicators of disengagement:

- Eye tracking
- Facial emotion recognition
- Other biometric indicators
- Completion of tasks required to complete more complex items
Other Ways CBTs Could Adapt

A CAT could, in principle, adapt to the presence of other construct-irrelevant factors such as:

- Test anxiety
- Cheating behavior
- Verbal ability
- Test time (if it is construct irrelevant)
Closing Thoughts

✚ We have only begun to explore ways in which CATs could be adaptive.

✚ We should be guided by a desire to maximize the validity of individual test scores.

✚ This implies that we should strive to reduce the effects of construct-irrelevant factors.

✚ It will require us to evolve toward more individualized, less standardized testing practices.
Thank You for Your Engagement

Questions?

steve.wise@nwea.org