

123



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.

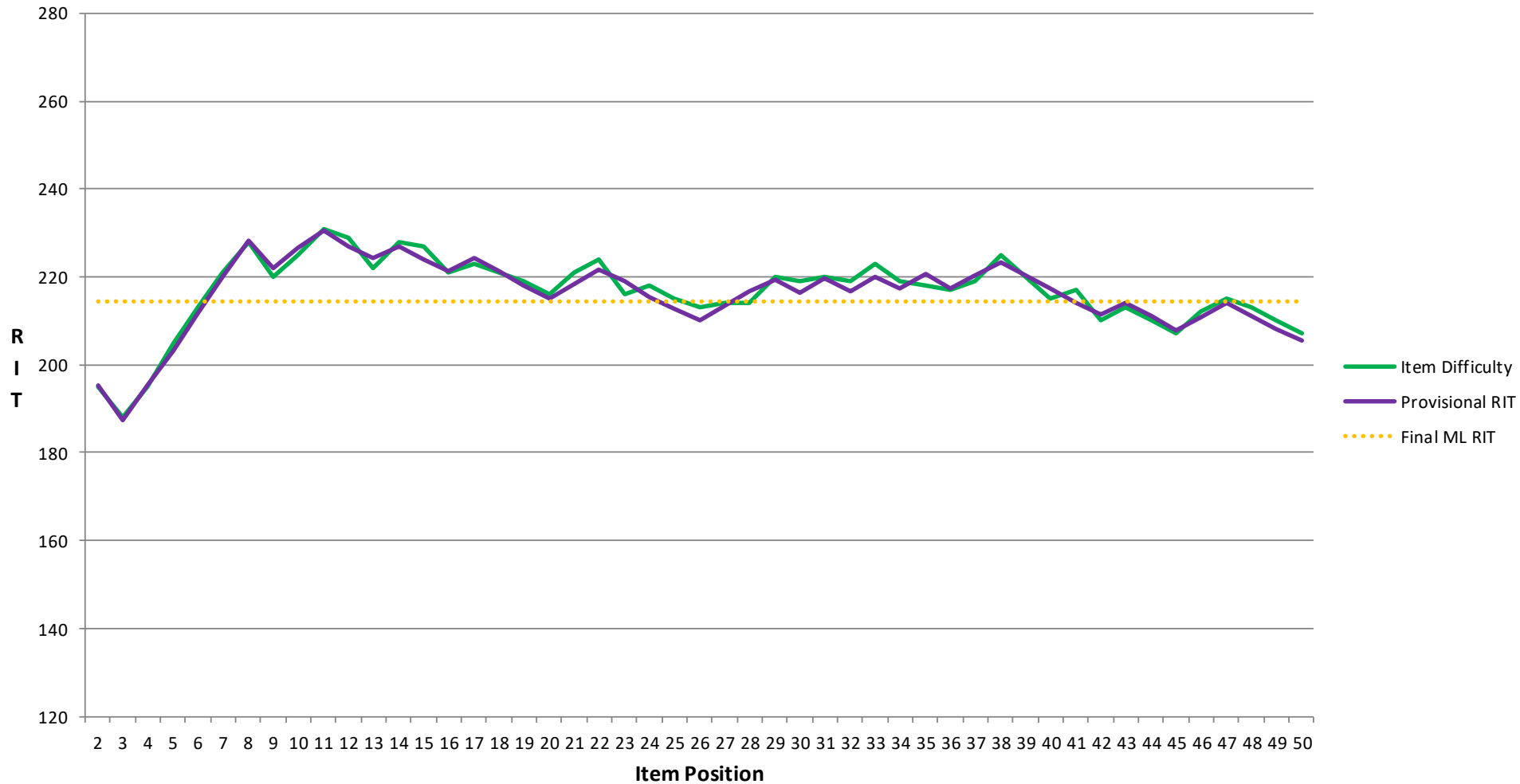
Experimental Results

- + 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.

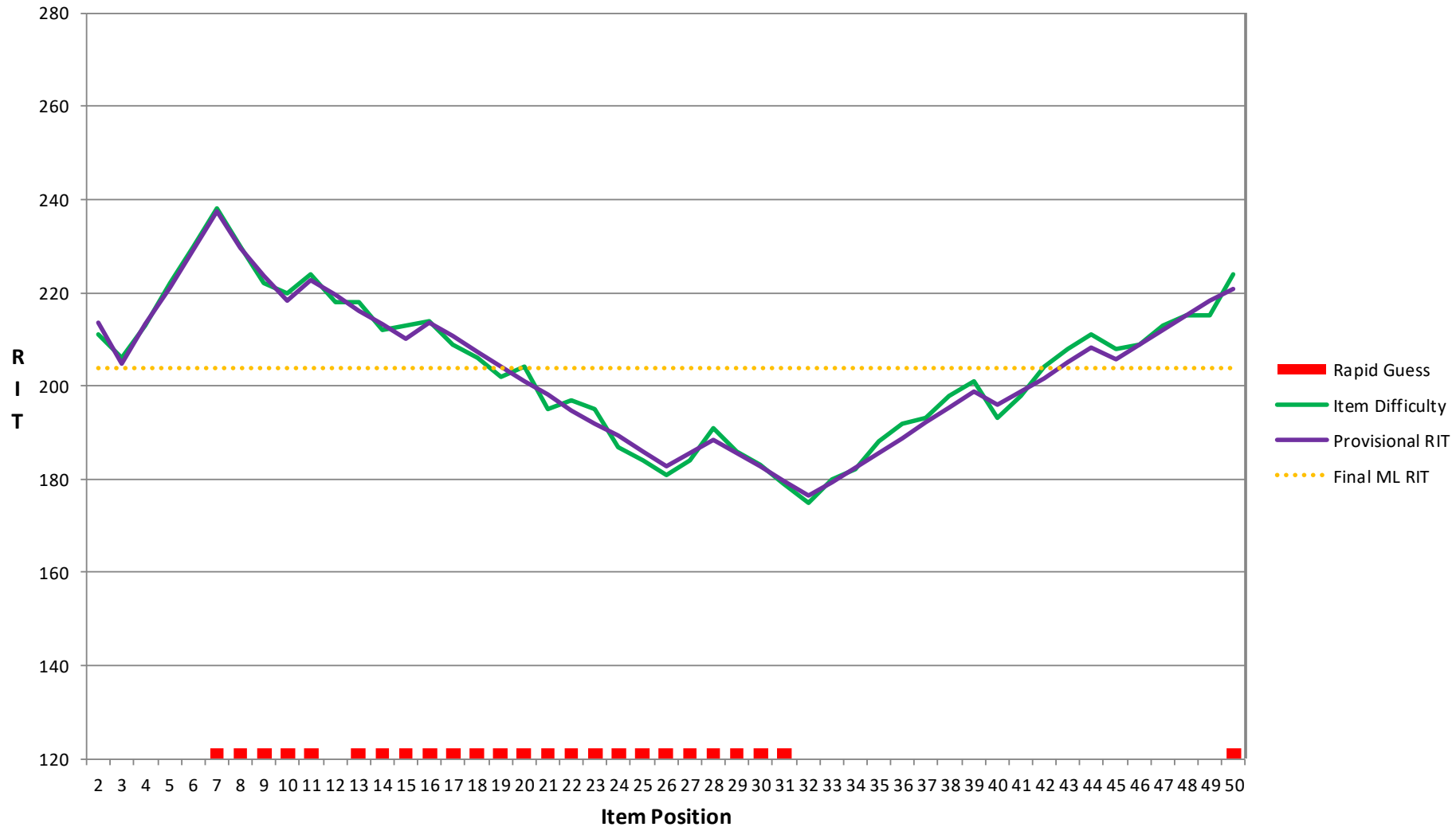
Disengagement in CATs

- + 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



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).



- IMPORT PROFILES
- MANAGE USERS
- MANAGE STUDENTS
- MANAGE TEST SESSIONS
- MAP SKILLS
- VIEW REPORTS ▾
- MODIFY PREFERENCES ▾

TEST STUDENTS

Testing Session Name: t05 Session Password: 7746

Proctor Interrupt PIN ⓘ

Total Students: 15
Testing: 9

Proctor Action Needed:
To Be Confirmed: 0
Paused: 0

Confirm Now

End Testing Session

Add More Students

Create Student

Refresh Status

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



Student(s) Disengaged ▾

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

Select Status... ▾

Select Action... ▾

Assign Test(s)

Assign Accommodations

Remove Student(s)

<input type="checkbox"/>	Last Name	First Name	Status	Approximate Question #	Test Assignment	Accommodations	Disengaged
<input type="checkbox"/>	Adams	Debbie	Testing	17	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Asimov	Cindy	Testing	15	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Barclay	Stan	Testing	8	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Cassidy	Tom	Testing	11	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Conner	Jane	Testing	7	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Douglas	Jim	Testing	8	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Heinlein	Terrie	Testing	12	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Harrington	Pat	Testing	9	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Hernandez	Tom	Testing	7	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Kobol	Samantha	Testing	9	Growth: Math 2-5 MS 2007	No	



- IMPORT PROFILES
- MANAGE USERS
- MANAGE STUDENTS
- MANAGE TEST SESSIONS
- MAP SKILLS
- VIEW REPORTS ▾
- MODIFY PREFERENCES ▾

TEST STUDENTS

Testing Session Name: t05 Session Password: 7746

Proctor Interrupt PIN ⓘ

Total Students: 15
Testing: 9

Proctor Action Needed:
To Be Confirmed: 0
Paused: 0

Confirm Now

End Testing Session

Add More Students

Create Student

Refresh Status

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

1 Student(s) Disengaged ▾

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

Select Status... ▾

Select Action... ▾

Assign Test(s)

Assign Accommodations

Remove Student(s)

<input type="checkbox"/>	Last Name ▲	First Name ◆	Status ◆	Approximate Question # ◆	Test Assignment ◆	Accommodations ◆	Disengaged ◆
<input type="checkbox"/>	Adams	Debbie	Testing	17	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Asimov	Cindy	Testing	15	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Barclay	Stan	Testing	8	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Cassidy	Tom	Testing	11	Growth: Math 2-5 MS 2007	No	⚠
<input type="checkbox"/>	Conner	Jane	Testing	7	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Douglas	Jim	Testing	8	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Heinlein	Terrie	Testing	12	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Harrington	Pat	Testing	9	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Hernandez	Tom	Testing	7	Growth: Math 2-5 MS 2007	No	
<input type="checkbox"/>	Kobol	Samantha	Testing	9	Growth: Math 2-5 MS 2007	No	



- IMPORT PROFILES
- MANAGE USERS
- MANAGE STUDENTS
- MANAGE TEST SESSIONS
- MAP SKILLS
- VIEW REPORTS ▾
- MODIFY PREFERENCES ▾

TEST STUDENTS

Testing Session Name: t05 Session Password: 7746

Proctor Interrupt PIN ⓘ

Total Students: 15
Testing: 9

Proctor Action Needed:
To Be Confirmed: 0
Paused: 0

Confirm Now

End Testing Session

Add More Students

Create Student

Refresh Status

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

1 Student(s) Disengaged ^

Apply a change to multiple students by choosing from the Select Status dropdown, then selecting the change to apply

Select Status...

Select Action...

Assign Test(s)

Assign Accommodations

Remove Student

<input type="checkbox"/>	Last Name ^	First Name ◆	Status ◆	Approximate Question # ◆	Test Assignment	Engaged ◆
<input type="checkbox"/>	Adams	Debbie	Testing	17	Growth: Math 2-5 MS 2007	Engaged
<input type="checkbox"/>	Asimov	Cindy	Testing	15	Growth: Math 2-5 MS 2007	Engaged
<input type="checkbox"/>	Barclay	Stan	Testing	8	Growth: Math 2-5 MS 2007	Engaged
<input type="checkbox"/>	Cassidy	Tom	Testing	11	Growth: Math 2-5 MS 2007	Disengaged
<input type="checkbox"/>	Conner	Jane	Testing	7	Growth: Math 2-5 MS 2007	No
<input type="checkbox"/>	Douglas	Jim	Testing	8	Growth: Math 2-5 MS 2007	No
<input type="checkbox"/>	Heinlein	Terrie	Testing	12	Growth: Math 2-5 MS 2007	No
<input type="checkbox"/>	Harrington	Pat	Testing	9	Growth: Math 2-5 MS 2007	No
<input type="checkbox"/>	Hernandez	Tom	Testing	7	Growth: Math 2-5 MS 2007	No
<input type="checkbox"/>	Kobol	Samantha	Testing	9	Growth: Math 2-5 MS 2007	No

Student Engagement tracks how fast questions are being answered and determines if Proctor intervention should be taken.

The 1 student below is skipping through test questions.

Encourage the student to make their best effort.

Refreshes every 20 seconds.

Cassidy, Tom ✓ ⚠ ✓
1 min ago

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

