MUTATED CATS: WHEN IS A CAT NOT A CAT?

- MERA Sp. 2016

MARK D. RECKASE

MICHIGAN STATE UNIVERSITY

THE IDEA OF AN ADAPTIVE TEST IS VERY OLD

- This was the standard way to evaluate students in the 1800s.
- The Binet intelligence tests from the early 1900s were adaptive tests.
- When I first started teaching at the university level, I used adaptive testing concepts to evaluate students.
  - I would ask a question and pick the next one based on the answer to the first one.
  - At that time, I thought I invented CAT. I was wrong.

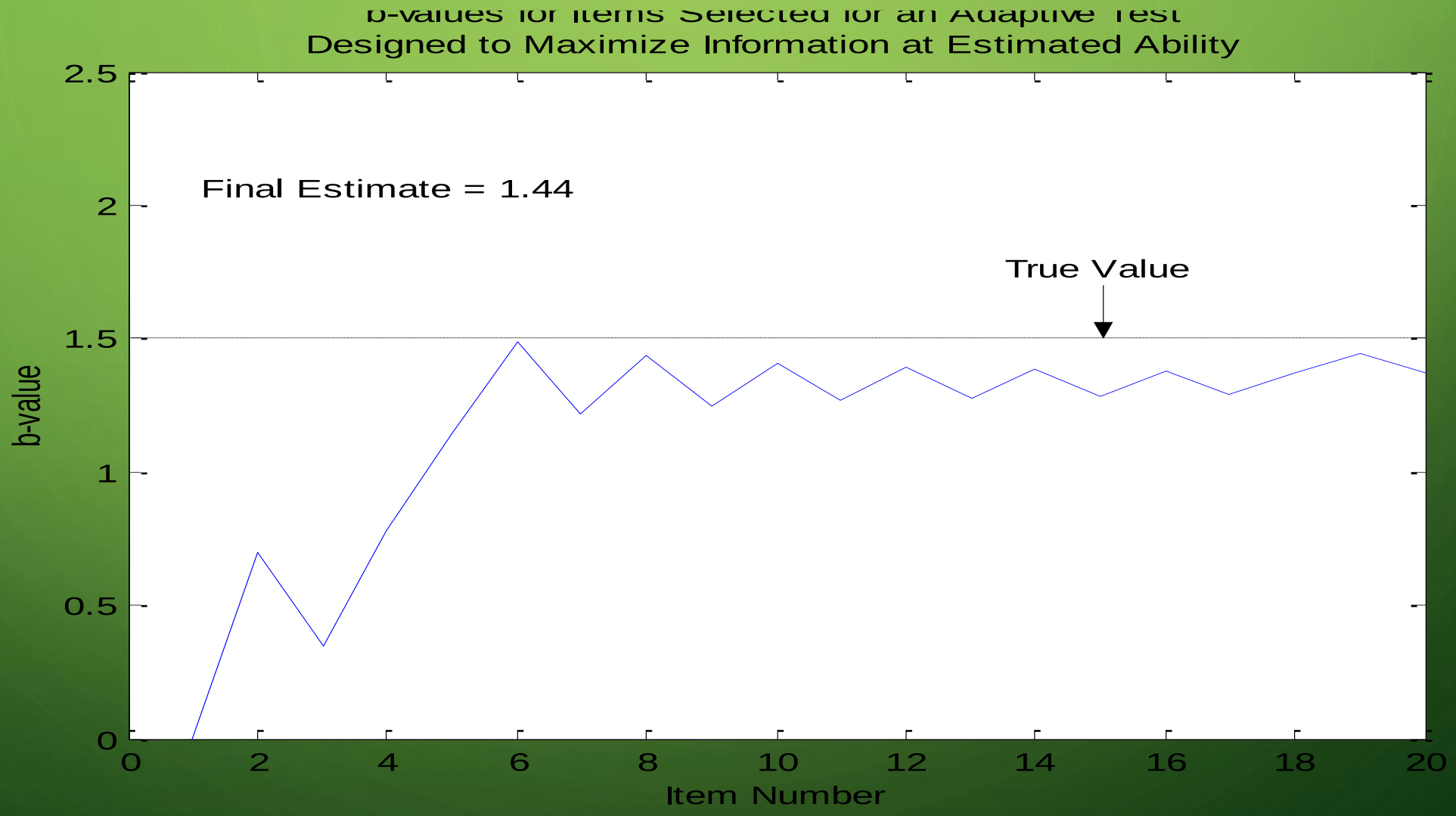
WHAT IS A CAT?

- The term CAT is an abbreviation for Computerized Adaptive Test
  - This term was coined by David Weiss of the University of Minnesota in the 1970s.
  - Alternative terms are tailored test, response contingent test, and pyramidal test.
  - A special case of a CAT is a multi-stage test.
- Initial research on CAT procedures showed that a CAT could give equal score precision to a fixed test with about ¼ the items.

THE WAY A CAT FUNCTIONS

- An initial estimate of proficiency is selected based on whatever information is available.
- An item is selected to best measure at the selected proficiency.
- After the item is administered, the estimate of proficiency is updated.
  - After a correct response the estimate of proficiency goes up.
  - After an incorrect response the estimate of proficiency goes down.
- The next item administered is selected to be best at measuring at the new estimate – more difficulty after a correct response, easier after incorrect.
- The process continues until the desired level of accuracy is achieved.

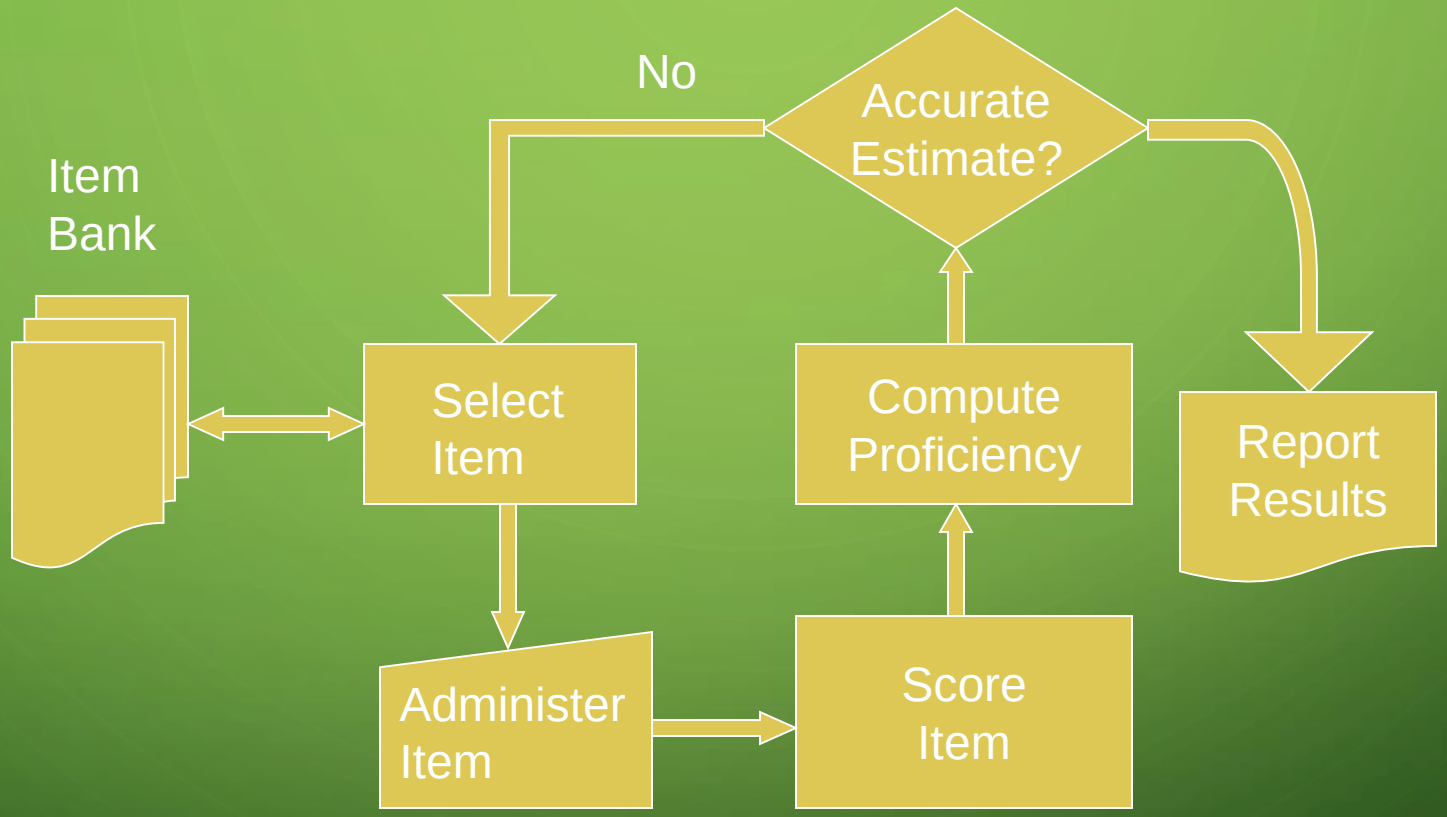
CONVERGENCE OF ESTIMATE OF PERFORMANCE



COMPONENT PARTS OF A CAT

- Calibrated Item Pool – To implement a CAT, computer storage with items and information about the item difficulties is needed.
- An Item Selection Algorithm – A criterion for selection items from the item pool based responses to items administered up to that point.
- A Score Estimation Procedure – A method is needed to estimate a point on the continuum defined by the calibration of items.
- A Stopping Rule – A method for deciding when the test is finished (often fixed length or variable length to fixed standard error).

COMPUTERIZED ADAPTIVE TESTING



REQUIREMENTS FOR A SIMPLE CAT

- Unidimensionality assumption for the item response theory model that is used.
- 100 to 200 items of varying difficulty
- A rule for selecting items to match proficiency – often maximum information
- A method for estimating the proficiency of an examinee
- A rule for deciding when to stop testing – often standard error of fixed length

CHALLENGES TO THE SIMPLE CAT

- Most achievement tests do not meet the unidimensionality assumption of the IRT models. It is assumed that the IRT models are robust to violations of unidimensionality.
- To minimize the cost of computers, want to administer the tests over a long administration window.
- Because of the long window, need to worry about test security – early examinees sharing with later examinees.

APPROACHES TO DEALING WITH THE CHALLENGES

- Violations of dimensionality assumption
  - Develop short CATs for each dimension in the test. This approach was used for the ASVAB and for ACT COMPASS.
  - Require content balancing to make sure all examinees get all content.
    - But content balancing violates the unidimensionality assumption.
    - It forces student to take items from content areas that might not be appropriate. This is required for fixed tests, but not adaptive tests.

APPROACHES TO DEALING WITH CHALLENGES

- Having a long testing window because not enough computers are available.
  - To make full use of computers, test several times every day for a period of time.
    - This raises security concerns because students can tell later students about items.
    - As a result, exposure control procedures have been developed. This requires larger item pools.
  - Might use a variety of different kinds of computers to make more available – desktops, notebooks, tablets.
    - Different computers have different screen sizes and resolution.
    - Different computers have different input devices – keyboards, etc.
    - Do these make a difference on the test?

UNINTENDED CONSEQUENCES OF MEETING CHALLENGES

- Item types/presentation may be restricted do to screen size and input device limitations. This is especially an issue of multiple types of computers are used.
  - This is especially an issue with items that have stimulus materials like reading tests and big science items.
- Content balancing reduces options for adaptation unless there is a very large item pool.
- Exposure control reduces adaptation unless there is a very large item pool.

MANAGING THE UNINTENDED CONSEQUENCES

- The unintended consequences are not a requirement of CAT.
  - Good interface design can allow presentation of most existing item types and many new ones are possible. But some standardization of hardware is needed.
  - The convenience of using the Internet for delivery has drawbacks because of lack of control and load issues.
  - Exposure control requires larger item pools. Good pool design is needed to keep the level of adaptation that is desired when exposure control is used.

MANAGING UNINTENDED CONSEQUENCES

- Content balancing is the biggest problem.
  - It requires violation of the IRT assumption of unidimensionality. If items are deleted to get better fit to the scoring model, content balancing is not really working.
  - It forces low scoring students to take hard items and high scoring students to take easy items reducing the advantage of an adaptive test. It is counter to adaptation.
  - It requires larger item pools to allow adaptation and exposure control.
- It would be better to have a number of short, unidimensional tests.

FINAL THOUGHTS

- I have no doubt that tests in the future will be delivered and scored by computer.
- But, like anything else, the procedures can be developed in a quality way or they can be done to keep costs down and to minimize hassle.
- I am working on ways to assess the amount of adaptation that actually takes place to inform users. My hope is this will push the field to solve the problems of unintended consequences.
- I want you to help me with that by being informed consumers.