

Computerized Adaptive Testing: Improving Assessment Through Technology

Barth Riley, Ph.D.

University of Illinois at Chicago
State of the Science Conference
May 31, 2006

The State of Health Outcomes Measurement

- The need for valid and reliable measures of a variety of health outcomes specific to persons with disabilities is growing.
- Many measures are available for the same construct, but:
 - Often disability or health condition specific.
 - Not comparable.
- The need for health outcomes assessment has resulted in an increase in ***respondent burden***, particularly for persons with disabilities.

Generalized vs. Targeted Measures

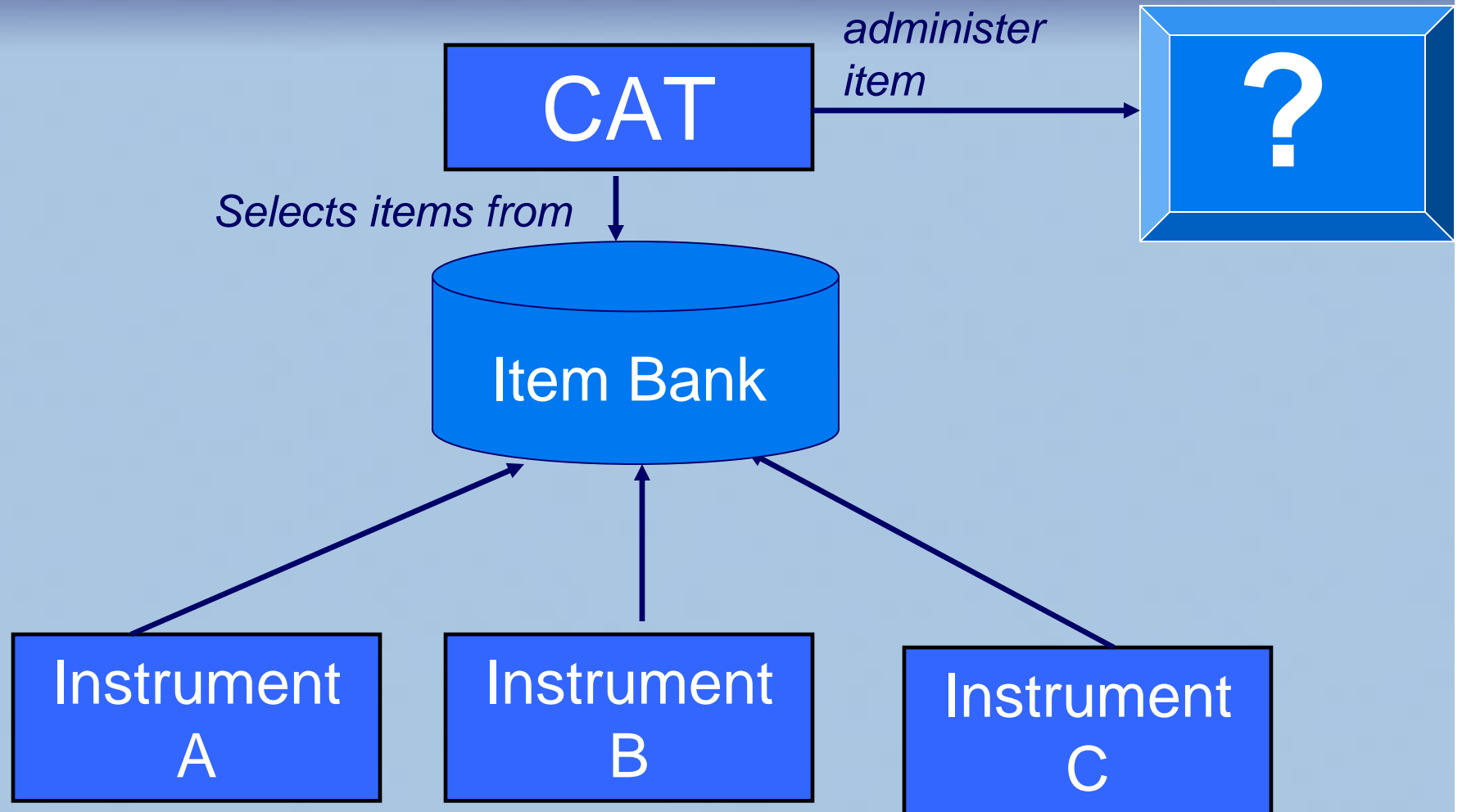
□ Generalized

- Heavy response burden
- Lack specificity

□ Targeted

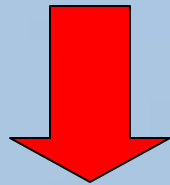
- Floor and ceiling effects
- Limited content validity
- Don't "talk with each other"

Tailoring Health Outcome Measurement

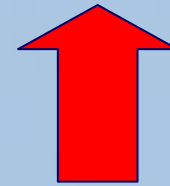


Benefits of CAT and Item Banking

CAT

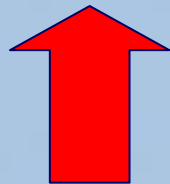


Respondent
Burden

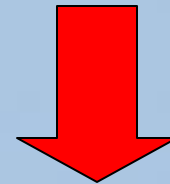


Tailoring/
Specificity

Item Bank

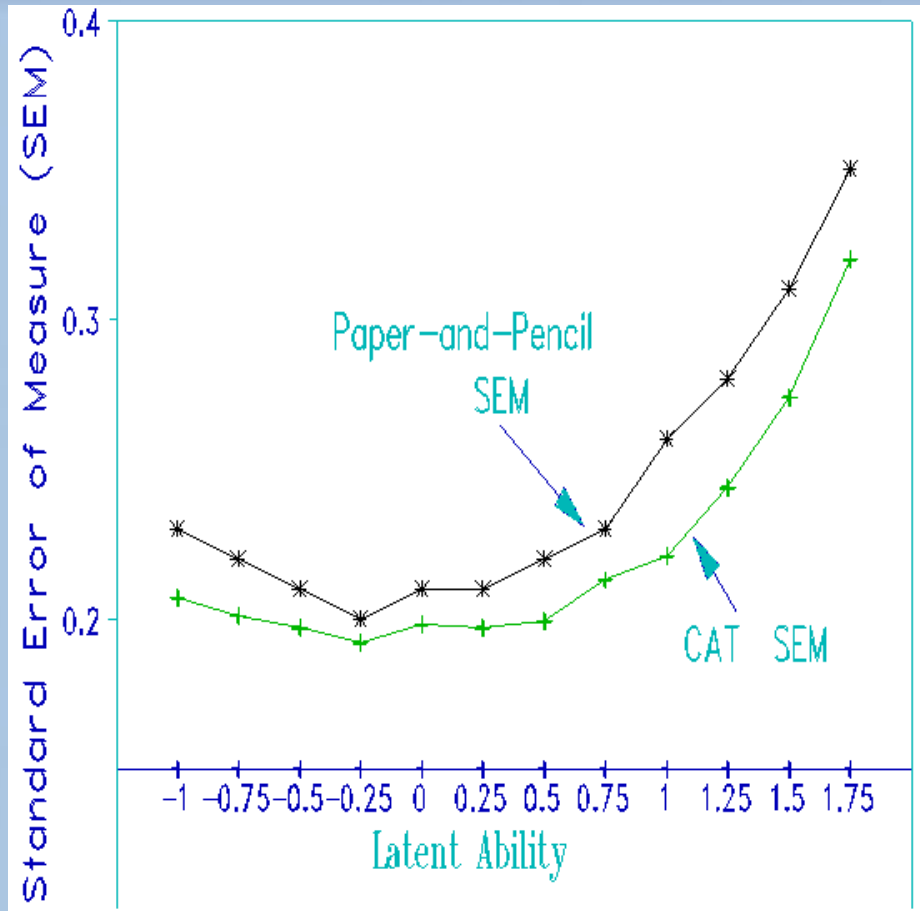


Coverage of
content
domains



Floor and
ceiling effects

CAT vs. Short Forms

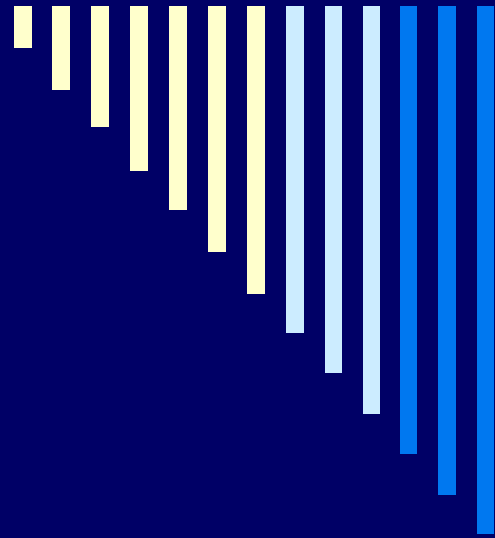


- CAT has been found to be superior to “short forms” of tests, yielding more precise measures.

CAT Example

□ AIMFREE Swimming Pool Survey

- Original survey consisted of 30 items.
- Rasch analyzed survey consists of 27 items.
- CAT version: 7 to 11 questions.



CAT: What Is It and How Does It Work?

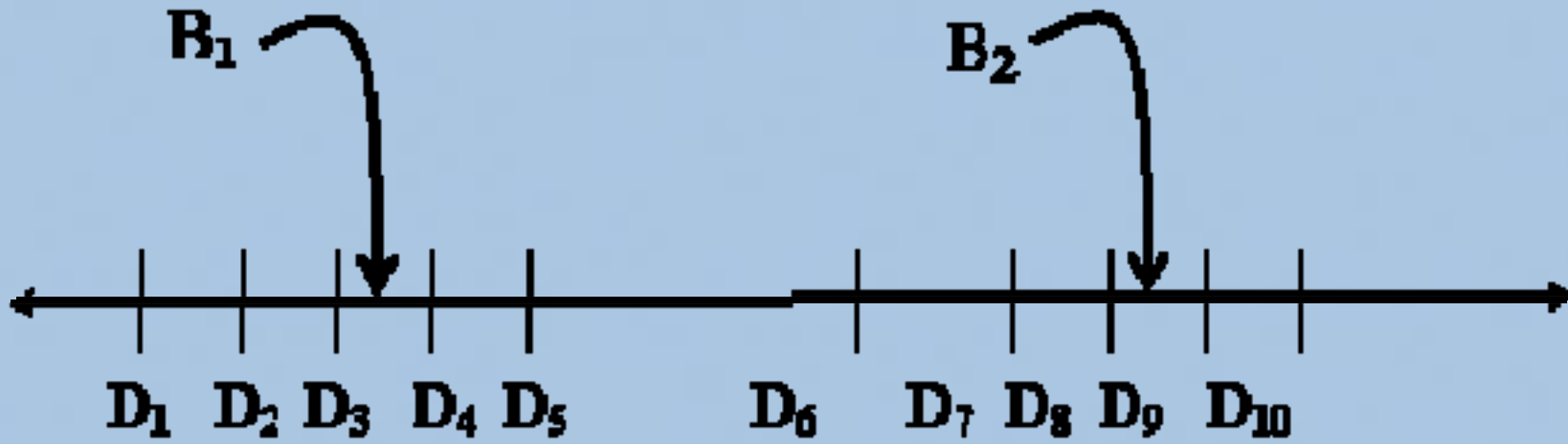
Item Selection

- There are several methods for selecting items during a CAT.
- The most common method is to find the item that provides the most information given the current estimate of the measure.
- The most informative item is usually the one which has a calibration (difficulty level) that most closely approximates the current estimate of the measure.
- Or, it is the item that the respondent has a 50% probability of passing or endorsing.

Item Selection cont.

Less Able Person

More Able Person



Easy Items

Hard Items

Item Selection cont.

- Item selection can also take into account the types of domains of items one wishes to have represented in each CAT administration.
 - Example: including items pertaining to strength, flexibility, and cardiovascular exercise in an adaptive physical activity measure.

Stop Rules

- The stop rule, which determines when the item administration process of the CAT ends, can be based on:
 - Measurement precision.
 - Number of items administered.
 - Test-taking time.
 - Some combination of the above.

Item Bank Size

- The more items there are in an item bank, the more likely it is that items that are tailored to an individual's level on the measured variable will be available.
- Typically, item banks consist of hundreds of items.
- The number of items will likely depend on
 - The number of constructs or domains being assessed.
 - Whether one wishes to estimate a measure or classify persons into groups.

Item Bank Blueprint

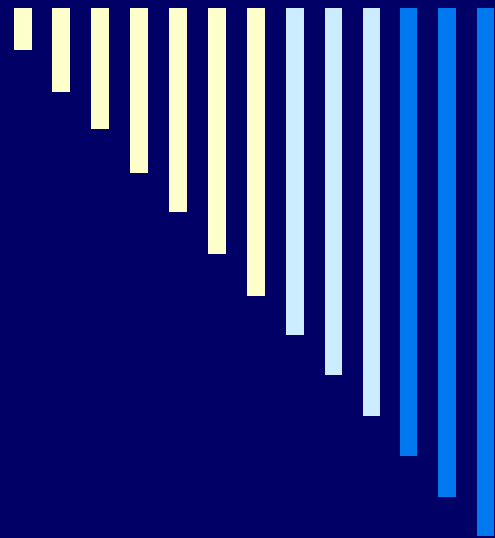
| | Cardio | Flex. | Strength | Household | Recreation | Sports | Swim. | Walk Wheel. |
|-----------|---------------------------|-------|----------|-----------|------------|--------|-------|-------------|
| Freq. | 21 | 18 | 22 | 33 | 38 | 30 | 7 | 15 |
| Duration | 21 | 18 | 22 | 33 | 38 | 30 | 7 | 15 |
| Intensity | 21 | 18 | 22 | 33 | 38 | 30 | 7 | 15 |
| Total | 187 activities, 561 items | | | | | | | |

Challenges and Considerations

- ❑ Survey responses can be affected by level of computer experience/anxiety.
- ❑ Persons with less familiarity with computers who are typically older and have less education are less likely to complete CAT-based surveys.
- ❑ Item bank development and maintenance is costly and time-intensive.
- ❑ Usability testing is needed to ensure accessibility and ease of use.

Accessibility Issues

- While computers can provide greater access to survey materials via large print and multimedia. Accessibility of the software (user interface) and hardware need to be considered.
 - Computers used for computerized assessment should be in a distraction-free setting.
 - Wheelchair accessible workstation.
 - Trackball and mouse support.
 - Compliance with Section 508, W3C, and Bobby standards.



CAT: Assessing the Recreation Environment

Environmental Assessment

- CAT, coupled with available wireless technology, can provide an efficient mechanism for evaluating the accessibility or recreation and other environments.
- To date, only one study has examined the use of wireless technology in conjunction with computerized adaptive testing.

Environmental Assessment



Assessment is done on the computer, with context sensitive help and photos stored with each item.



Team uses computer-based version of AIMFREE on Tablet PC with a USB camera.



Broadband allows real-time remote viewing and guided image capture..

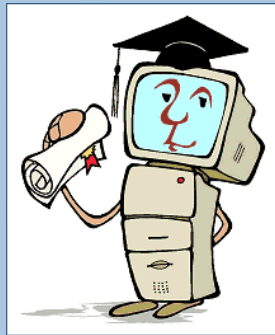


Completed assessment is uploaded and shared electronically with the TRT



TRT recommendations report is sent to assessment team and facility. Solutions are cataloged in a database.

Remote Assessment: The Future



Computerized adaptive testing (CAT) and integrated *smart tools* assure a valid survey with a minimum number of items



Solutions wizard matches items for specific access issues with possible solutions from the cumulative knowledge base and sends them to the team.

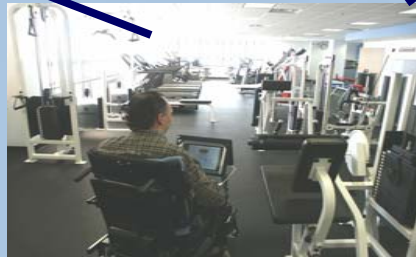
Items without a good match

Or items for which the team feels the existing solutions are not adequate or appropriate.

Review by the TRT with report sent to survey team and facility

New solutions get added to the solutions knowledge base

Ubiquitous broadband allows survey teams to access Web application from anywhere.



Issues in Remote Environmental Assessment

- Reliability and availability of the technology
- How should items be selected?
 - Based on responses to previous items (standard approach).
 - Based on respondent's current location at the site (e.g., equipment room, locker room).
 - Which items are needed to generate solutions?