Advantages of *Computerized Adaptive Testing (CAT)*

Nathan A. Thompson, PhD

White Paper - August, 2011

ASSESSMENT SYSTEMS
6053 Hudson Road, Suite 345
St. Paul, MN 55125 USA

Computerized adaptive testing (CAT) is a next-generation method of delivering tests to students. CAT works by tailoring both the **difficulty** and **number** of items to an examinee; smarter examinees warrant more difficult items. Designing the test with this interactive, intelligent CAT algorithm has numerous benefits, and this paper discusses some of those benefits.

Problems with traditional tests

In a traditional paradigm of test design, tests are assembled into one or more fixed forms. We have all experienced such tests: *here are 200 multiple-choice items, please answer them all*. This approach has a number of problems, some obvious and some more subtle.

The most obvious disadvantage is that this approach is quite inefficient. All 200 items are not necessary for all examinees; if someone were to take only the 50 most difficult items and answer 45 correctly, we can confidently assign a high score without wasting the time to administer the 150 easy questions. Doing so not only lightens the burden for the student, but can also translate to substantial savings in testing costs.

Another issue is the large differences in precision among examinees. This is due to traditional tests typically being built with large numbers of medium-difficulty items. There is good reason for this: we have large numbers of medium-ability people! Plus, classical reliability is highest in this situation. But this is at the expense of low or high ability students. Those examinees are measured with much less precision.

For the same reason, low and high ability examinees typically have a poor experience with the test. Low examinees are exhausted and discouraged by the fact that most items are too difficult to be relevant. Meanwhile, high examinees are not appropriately challenged, because most items are too easy.

The fixed-form approach has a number of additional drawbacks when classical test theory (CTT) is used. A well-known issue is the effect of guessing, whether random or nonrandom. There are classical methods to correct for guessing, but they are actually more biased than not using any correction at all. Most CATs are based on item response theory (IRT), which provides a model-based approach to account for guessing. Additionally, IRT provides much more powerful procedures for linking and equating, ensuring a stable scale with comparable scores for all examinees.

What is CAT?

CAT works by tailoring both the **difficulty** and **number** of items to an examinee. Adapting the difficulty of items means that an examinee with high ability will receive difficult items, while an examinee with low ability will receive easy items. Adapting the number of items means that the test will stop when a certain psychometric criterion is met; some examinees will need more items than others, meaning that the test can finish quickly for some examinees. Both of these adaptive aspects lend to the advantages discussed below.

Want to learn more about how CAT works? A separate white paper delves into this topic more deeply. It provides a background on IRT and then describes each of the algorithms necessary to deliver a CAT.

Some Advantages of CAT

Shorter tests: Research (e.g., Weiss & Kingsbury, 1984) has shown that CAT can reduce testing time by 50% or more. This can obviously translate into huge financial benefits. Suppose an organization tests 50,000 candidates per year, paying $50 per hour for seat time at professional

centers. If the test is decreased from 2 hours to 1 hour, the annual costs are reduced by $250,000.

Equiprecision: CATs can be designed so that examinees are all measured with the same level of precision, even though they all potentially see different items. This makes the test extremely fair from a psychometric perspective.

More precision: Because CATs are more efficient, the organization has the option to design the CAT to actually be more precise than a conventional test while still using fewer items.

Examinee experience: As mentioned previously, a CAT will provide an appropriate challenge for each examinee. Low examinees are not discouraged or intimidated. High examinees enjoy receiving difficult items.

Increased motivation: Because of the better experience, there is likely an increase in examinee motivation. Low examinees feel better, and high examinees feel challenged. Both will try harder than with a conventional test.

Security: Because the CAT algorithm is very flexible and can adapt with potentially millions of permutations, there is much greater security than if everyone was administered the same set of 200 items.

Frequent retesting: The high number of permutations also enables more frequent retesting. If a student takes a CAT, then receives a few weeks of instruction, by the time they take a CAT again their ability has increased somewhat, and they will receive a completely different test.

CBT advantages: CAT also benefits from the same advantages of non-adaptive computer-based testing (CBT). For example, tests delivered by computer (whether adaptive or not) can easily utilize multimedia such as audio and video files. This topic is also discussed in a separate white paper.

Disadvantages of CAT

No review: CATs rarely allow for examinees to return to items already administered, as the CAT has since adapted and it cannot unadapt.

Item exposure: CATs are designed to select the best items in the bank, and these items will often become overexposed if a control algorithm is not implemented.

Recovery of poor starts: CATs are susceptible to issues with examinee test anxiety, as the elimination of item review prevents someone from going back to the first few items. If they answered all those items incorrectly due to severe test anxiety, the test cannot correct itself.

Public relations: Because of the complexity and the departure from the familiarity of the traditional exam paradigm, an organization must put forth more effort into public relations, explaining CAT and the reasons for using it.

Requires calibration: Nonstandard item formats are often not easily scored in real time, or are not able to be calibrated with IRT. Both are required for CAT.

Requirements: CAT requires large sample sizes and extensive expertise. Testing programs need at least several hundred examinees, and an experienced Ph.D. psychometrician. Another white paper discusses the requirements to develop a CAT examination program, but some information is presented next.

How FastTest can help implement

Successful implementation of a CAT exam requires two facets: **software** and **expertise**.

First, a test delivery system but be established that is capable of fully adaptive tests based on IRT. *FastTest* is the **only content-free system for online testing that provides fully adaptive CAT**. Other content-free systems might claim to provide adaptive capabilities, but are based on crude approximations such as fixed branching.

Additionally, a psychometrician is necessary to perform the quantitative analysis and extensive research needed for CAT. Publishing a CAT without such an expert will likely make it inefficient, indefensible, or both, seriously hampering validity. *FastTest* can also provide this essential expertise, from one of our four Ph.D. psychometricians, and at reasonable rates.

About *FastTest*

FastTest is a comprehensive web-based system for the entire test development cycle:

- Item banking
- Item review
- Test assembly
- Standard setting
- Test delivery (computer or paper)
- Adaptive testing (CAT)
- Score reporting
- Results management
- Back-end reporting
- Psychometric analysis.

While comprehensive, it is flexible and scalable enough to be applied to testing organizations of all types and sizes. Custom solutions are also available.

For more information, visit www.assess.com or contact sales@assess.com

ASSESSMENT SYSTEMS