

AS SEEN IN DEANS DIGEST

Demystifying the GMAT: Computer Adaptive Testing

By Lawrence M. Rudner

Rather than using the computer as an electronic page turner, the GMAT exam uses the computer's processing power to analyze each examinee's responses during the test session. By having the computer calculate a final score estimate after each question and using that estimate to select subsequent questions, GMAC is able to provide a more valid, reliable, secure, and shorter test. Here's a look at the logic and advantages of computer adaptive testing.

Logic

An individual's test begins with a randomly selected question of average difficulty, drawn from a large pool of test questions. Subsequent questions are then selected from the pool with the following basic steps:

1. The examinee responds to the question.
2. The computer estimates the examinee's final score from his responses and the difficulty of the limited number of questions he has received. Correct responses to relatively hard questions will result in higher estimated scores. Incorrect responses to relatively easy questions will result in lower estimated scores.
3. The computer then evaluates all eligible questions covering the necessary content to determine which will be the most informative questions to administer next, given the examinee's current estimated score.
4. One of the best "next questions" is administered next. Typically, the best "next questions" will be relatively harder as the estimated score gets higher, and relatively easier as the estimated score gets lower.
5. Steps 1 through 4 are repeated until the required number of questions has been administered to ensure test accuracy and reliability.

By estimating the examinee's final score after each response, the computer tailors the test based on both the difficulty of the previously administered questions and the examinee's responses. With the right pool of test questions and the right question selection algorithm, CAT can be much more efficient than a traditional, fixed-question test in which all examinees answer the same set of questions. Such a general-purpose test must have questions spanning the entire score range. As a result, the test must be longer, and every test taker sees some questions that are much too difficult and some that are much too easy.

Advantages

The key requirements for a quality CAT program are a sizeable collection of quality test questions, a careful approach to determining which questions are out in the field at any one time, a question selection algorithm that assures every individual receives an equal mix of test content and one of the best "next questions," and an active security component because test questions are re-used. When these requirements are met, a CAT program can realize numerous advantages for the test taker and in quality:

- Significantly less time is needed to administer CATs than fixed-question tests. Fewer questions are needed to achieve acceptable ac-

curacy. CATs can reduce testing time by more than 50 percent while maintaining the same level of reliability.

- Shorter testing times also reduce fatigue, a factor that can significantly affect an examinee's test results.
- Tests are individually paced, so an examinee does not have to wait for others to finish before going on to the next section.
- Because the questions are tailored to the examinee, there is less likelihood of error due to lucky guesses or missing a question that should not have been missed.
- Tests can be given "on demand," and scores can be available immediately.
- CATs can provide accurate scores over a wide range of abilities, including the top test takers, while traditional tests are most accurate for a narrow range of average ability examinees.
- Test administrator differences are eliminated as a factor in measurement error.
- Test security is increased because hard copy test booklets are never compromised, and it is very rare for two examinees to see the same set of questions.

Computer adaptive testing is the state-of-the-art for assessment in many fields and is the mode used by many well-known and well-run testing programs, such as the National Institute for Educational Measurement in the Netherlands (CITO), the Armed Services Vocational Aptitude Test Battery (ASVAB), the Uniform Certified Public Accountant examination (AICPA), COMPASS from ACT, the Microsoft Certified Professional exams, the North American Pharmacist Licensure Examination (NAPLEX), the National Council Licensure Examinations (NCLEX), and Renaissance Learning Corporation's STAR tests (K-12). GMAC researchers have been active in this field and are recognized as the world leaders.

To learn more about computer adaptive testing and to play with a CAT, see my online interactive tutorial at <http://echo.edres.org:8080/scripts/cat/catdemo.htm>.

Lawrence M. Rudner, PhD, MBA, is vice president of research and development and chief psychometrician at the Graduate Management Admission Council. He can be reached at lrudner@gmac.com.

This article was published in the June 2010 issue of *Deans Digest*. Go to <http://www.gmac.com/gmac/NewsandEvents/DeansDigest/2010/June2010/Demystifying-the-GMAT-Computer-Adaptive-Testing.htm>

© 2011 Graduate Management Admission Council. All rights reserved. This article may be reproduced in its entirety without edits with attribution to the Graduate Management Admission Council.

GMAC®, GMAC logo, GMAT®, GMAT logo, and Graduate Management Admission Council® are registered trademarks of the Graduate Management Admission Council in the United States and other countries. www.gmac.com/trademarks