Graduate Management Admission Council®

# Some Thoughts on Controlling Item Exposure in Adaptive Testing


**Charles Lewis**

**Fordham University**

**and**

**Educational Testing Service**


*Keynote Address Presented June 7, 2007*

## Abstract

The issue of item exposure control for CATs is considered and critically evaluated relative to the primary goal of adaptive testing: providing valid and reliable measurement with a minimal number of items. It is also evaluated relative to problems of test security that the method is intended to address. The point is made that methods of exposure control should be developed and evaluated in the context of other test security measures and not in isolation. A simplified variation of a standard procedure for controlling item exposure rates, due to Sympson and Hetter (and further developed by Martha Stocking and the author) is introduced. A "toy" example of a CAT (a four-tem test with a 20-item bank) is considered and the simplified exposure control method is applied to illustrate how such a procedure functions in practice.

## Acknowledgment

**Presentation of this paper at the 2007 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.**



## Citation

**Lewis, C. Some thoughts on controlling item exposure in adaptive testing. (2007). In D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. Retrieved [date] from** www.psych.umn.edu/psylabs/CATCentral/

## Author Contact

**Charles Lewis, Psychology Department, Fordham University, Bronx, NY 10458**
clewis@fordham.edu

# Dedication

This paper is dedicated to my friend and colleague Martha Stocking, who passed away at the end of last year after a lengthy illness. Among her many accomplishments in the field of psychometrics, Martha made a number of theoretical contributions that were critical for the implementation of computerized adaptive testing in several major educational testing programs, including the Graduate Management Admission Test®. Among these contributions was the development of algorithms to attain specified rates of conditional item exposure control for adaptive tests (Stocking & Lewis, 1998, 2000). For an appreciation of her life and work, see Eignor (2007).

Martha L. Stocking
1942 – 2006

# Some Thoughts on Controlling Item Exposure in Adaptive Testing

For a recent comprehensive review of research on item exposure control for adaptive tests, see Georiadou, Triantafillou, and Economides (2007). This paper will make no attempt to duplicate their valuable work. Instead, it begins with a consideration of the context within which exposure control functions. As Davey and Parshall (1995) have pointed out, test security for a computerized adaptive test (CAT) (including control of item exposure rates) should not come at the cost of its primary goal: to provide valid and reliable measurement with a minimal number of items. This is also a theme of a recent NCME symposium on test security and CATs (Chang & Davey, 2007). In addition, this symposium broadened the discussion of test security beyond item exposure control to item design, item bank design, bank rotation or adjustment, and item "recycling." To this list, it is appropriate to add the continuous monitoring of items and test takers, with the goal of identifying items that might have been compromised for some examinees as well as examinees who might have had prior access to some items. In this broader context, controlling item exposure rates can be seen as assuming secondary importance for CAT security, analogous to using scrambled forms when administering paper-and-pencil tests. Moreover, it is important to note that the development and evaluation of any strategy for controlling item exposure rates must be evaluated in the context of the entire range of test security measures, and not in isolation.

## Illustrating Exposure Control

Sympson and Hetter's (1985) basic idea for controlling the rate at which items are used was to introduce a constant ($p_i$) to be specified for each item *i* in an item bank. If that item were selected at any point during a CAT, it would actually be administered with probability $p_i$ or dropped from the list of available items (with probability $1-p_i$) for that test. If the item were not administered, then the selection would proceed among the items still available for that test, each time determining the actual administration based on the probability $p_i$ for that item.

In the procedure described by Sympson and Hetter (1985), the conditional item administration probabilities (given selection) $p_i$ are determined iteratively, based on simulations of CAT administrations, with the goal that the maximum probability with which an item in the bank is administered be less than some chosen target value *r*. In this sense, the procedure is said to control the maximum item exposure rate for the CAT. Papers by Stocking and Lewis (1998, 2000) extended this idea to provide control of the maximum item exposure rate conditional on ability. Subsequent research, described by Georiadou et al. (2007), has generally supported the effectiveness of the Sympson-Hetter (S-H) approach in achieving its goal while maintaining the validity and reliability of the resulting CAT, and only marginally increasing test length.

To illustrate the basic ideas of item exposure control, a simplified, non-iterative version of the S-H procedure will be described. Specifically, let *r* be the desired maximum item exposure rate. Suppose that all values of $p_i$ are initially set equal to *r*. (For any S-H procedure, *r* will actually be a lower bound for each $p_i$.) Next let *g* be a random variable representing the number of items selected at each stage of testing until one is administered. Its distribution is given by

$$\Pr(g=1)=r, \tag{1}$$

$$\Pr(g=2)=r(1-r), \tag{2}$$

etc. Generally, its probability distribution function $(\text{for } g=1, 2, ...)$ is

$$p(g)=r(1-r)^{g-1}. \tag{3}$$

(It should be noted that $g-1$ has a Geometric distribution. For the present purposes, it is more convenient to work with the distribution of *g*.)

As a practical matter, all CAT item banks are finite. Thus, there is the danger with any S-H procedure of exhausting the bank of items before a given test length *n* has been reached. Sympson and Hetter (1985) addressed this problem by always setting at least *n* values of $p_i$ equal to 1 (so that at least *n* items would always be administered if they were selected). Stocking and Lewis (1998) proposed an alternative strategy that they referred to as the multinomial method. In the current context, this method involves truncating the distribution of *g*. This truncation should occur at a value no greater than the ratio of the number of items in the CAT bank to the number of items in the CAT, chosen so that the bank will never be exhausted before the complete test is administered. Table 1 illustrates this process for $r=.25$ and using a truncation value of $m=5$.

**Table 1. Probability and Cumulative Distributions for *g*, the Number of Items Selected Until One is Administered, Before and After Truncation (Based on $r=.25$)**

| | Original | | Truncated | |
|---|---|---|---|---|
| $g$ | $p(g)$ | $P(g)$ | $p_t(g)$ | $P_t(g)$ |
| 1 | .25 | .25 | .33 | .33 |
| 2 | .19 | .44 | .25 | .57 |
| 3 | .14 | .58 | .18 | .76 |
| 4 | .11 | .68 | .14 | .90 |
| 5 | .08 | .76 | .10 | 1.00 |

Stocking and Lewis (1998) advocated using the cumulative distribution function $P_t(g)$ of the truncated random variable *g* as follows: At each stage of the CAT, make an ordered list of the first *m* available items based on an item selection algorithm. Next select a random number *u* uniformly distributed on $(0,1)$. Let *g** be the smallest value of *g* such that $u<P_t(g)$. Drop the first $g^*-1$ items from the list of available items and administer item *g**. (Note that, in Table 1, the first item selected will have a probability of being administered equal to .33, rather than the initial value of .25.)

Now apply this exposure control rule to a "toy" example to see how such a simplified S-H procedure works. For a CAT consisting of four items, imagine a bank of 20 Rasch items (so the ratio of bank size to test length is $m = 20/4 = 5$, as in Table 1) with equally spaced difficulty parameters:

$$\{-1.9, -1.7, -1.5, -1.3, -1.1, -.9, -.7, -.5, -.3, -.1,\ .1,\ .3,\ .5,\ .7,\ .9,\ 1.1,\ 1.3,\ 1.5,\ 1.7,\ 1.9\}.$$

Take a prior density function for $\theta$ given by

$$p(\theta) \propto \left[\frac{\exp(\theta+2)}{1+\exp(\theta+2)}\right]\left[\frac{1}{1+\exp(\theta-2)}\right]. \tag{4}$$

This choice of prior can be thought of as the result of combining a "flat" prior with a likelihood corresponding to a positive response to an easy Rasch item with a difficulty parameter $b = -2.0$ and a negative response to a difficult Rasch item, with a difficulty parameter $b = +2.0$.

Next, define a preliminary estimate of $\theta$ after administering *i* items as the mode, denoted here by $\hat{\theta}_{(i)}$, of the posterior density $p\left(\theta \middle| \mathbf{y}_{(i)}\right)$, where $\mathbf{y}_{(i)}$ denotes the vector of responses to the *i* items already administered. The item selection algorithm to be used in this example selects the next item (*j*) to be the available item in the bank for which $\left|\hat{\theta}_{(i)} - b_j\right|$ is minimized. (In case of a tie among two or more items, suppose the algorithm selects the easier item.) To start, note that $\hat{\theta}_{(0)} = 0.0$. Table 2 gives the ordered difficulties of the "best" five items, along with the truncated distribution of *g*.

**Table 2. List Used to Select the First CAT Item**

| $g$ | $p_t(g)$ | $P_t(g)$ | $b_g$ |
|---|---|---|---|
| 1 | .33 | .33 | -.1 |
| 2 | .25 | .57 | .1 |
| 3 | .18 | .76 | -.3 |
| 4 | .14 | .90 | .3 |
| 5 | .10 | 1.00 | -.5 |

Now sample a uniform random number $u_{(1)} = .9902$. This implies $g^* = 5$, so the first administered item has $b_{(1)} = -.5$. This item, as well as the previous four items in the list in Table 1, are all removed from the original bank, giving the following reduced bank:

$$\{-1.9, -1.7, -1.5, -1.3, -1.1, -.9, -.7,\quad .5,\ .7,\ .9,\ 1.1,\ 1.3,\ 1.5,\ 1.7,\ 1.9\}.$$

Suppose $y_{(1)} = 0$. (The examinee gives a negative response to the first item.) The resulting posterior mode is $\hat{\theta}_{(1)} = -1.33$ and the new list of "best" available items is given in Table 3.

**Table 3. List Used to Select the Second CAT Item**

| $G$ | $p_t(g)$ | $P_t(g)$ | $b_g$ |
|---|---|---|---|
| 1 | .33 | .33 | -1.3 |
| 2 | .25 | .57 | -1.5 |
| 3 | .18 | .76 | -1.1 |
| 4 | .14 | .90 | -1.7 |
| 5 | .10 | 1.00 | -.9 |

Now a second uniform random number is sampled: $u_{(2)} = .5343$. The resulting $g^*$ = 2, so the second administered item has $b_{(2)} = -1.5$. This item and the item with $b = -1.3$ are removed to give the bank available for selecting the third CAT item:

$\{-1.9, -1.7, \quad -1.1, -.9, -.7, \quad .5,\ .7,\ .9,\ 1.1,\ 1.3,\ 1.5,\ 1.7,\ 1.9\}$.

Suppose $y_{(2)} = 1$. Then $\hat{\theta}_{(2)} = -.68$. The resulting list of best available items is given in Table 4.

**Table 4. List Used to the Select Third CAT Item**

| $g$ | $p_t(g)$ | $P_t(g)$ | $b_g$ |
|---|---|---|---|
| 1 | .33 | .33 | -.7 |
| 2 | .25 | .57 | -.9 |
| 3 | .18 | .76 | -1.1 |
| 4 | .14 | .90 | -1.7 |
| 5 | .10 | 1.00 | .5 |

To select the third CAT item, sample $u_{(3)} = .3998$, giving $g^*$ = 2 and $b_{(3)} = -.9$. After removing this item and the one with $b = -.7$, the available item bank for selecting the fourth CAT item is given by:

$\{-1.9, -1.7, \quad -1.1, \quad .5,\ .7,\ .9,\ 1.1,\ 1.3,\ 1.5,\ 1.7,\ 1.9\}$.

If $y_{(3)} = 0$, then $\hat{\theta}_{(3)} = -1.26$, and the fourth (and last) item list is given in Table 5.

**Table 5. List Used to Select the Fourth CAT Item**

| $g$ | $p_t(g)$ | $P_t(g)$ | $b_g$ |
|---|---|---|---|
| 1 | .33 | .33 | -1.1 |
| 2 | .25 | .57 | -1.7 |
| 3 | .18 | .76 | -1.9 |
| 4 | .14 | .90 | .5 |
| 5 | .10 | 1.00 | .7 |

Sample $u_{(4)} = .9124$, giving $g^* = 5$ and $b_{(4)} = .7$. If the final response $y_{(4)} = 0$, then $\hat{\theta}_{(4)} = -1.38$. The results are summarized in Table 6.

**Table 6. Summary of CAT Constructed Using S-H Procedure**

| $i$ | $u_{(i)}$ | $g^*_{(i)}$ | $b_{(i)}$ | $y_{(i)}$ | $\hat{\theta}_{(i)}$ |
|---|---|---|---|---|---|
| 1 | .9902 | 5 | -.5 | 0 | -1.33 |
| 2 | .5343 | 2 | -1.5 | 1 | -.68 |
| 3 | .3998 | 2 | -.9 | 0 | -1.26 |
| 4 | .9124 | 5 | .7 | 0 | -1.38 |

In this example, the maximum item exposure rate is .33 (vs. .20 for random selection in a non-adaptive four-item test). Note that this is a conditional, as well as an unconditional rate, for any conditioning variables, including $\theta$. It is also worth emphasizing that this simplified procedure might have unacceptable consequences in terms of validity, reliability and efficiency for actual CATs.

## Take-Home Message

Item exposure control has a modest role to play in the broader context of CAT design and security. Any procedure to control item exposure should be developed and evaluated in this broader context.

## References

Chang, H., & Davey, T. (Organizers). (2007, April). *The challenge of computerized adaptive testing: Increasing security without sacrificing efficiency*. Symposium conducted at the annual meeting of the National Council on Measurement in Education (NCME), Chicago, IL.

Davey, T., & Parshall, C. G. (1995, April). *New algorithms for item selection and exposure control with computerized adaptive testing.* Paper presented at the annual meeting of the American Educational Research Association (AERA), San Francisco, CA.

Eignor, D. (2007). In memoriam: Martha L. Stocking. *Applied Psychological Measurement, 31,* 167-168.

Georiadou, E., Triantafillou, E., & Economides, A. (2007). A review of item exposure control strategies of computerized adaptive testing developed from 1983 to 2005. *Journal of Technology, Learning, and Assessment, 5 (8).* Retrieved June 5, 2007, from http://www.jtla.org.

Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics, 23,* 57-75.

Stocking, M. L., & Lewis, C. (2000). Methods of controlling the exposure of items in CAT. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 163-182). Hingham, MA: Kluwer.

Sympson, J. B., & Hetter, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Military Testing Association,* (pp. 973-977). San Diego, CA: Navy Personnel Research and Development Center.