

■ **A Formal Characterization of and Some
Alternatives to Simpson-Hetter Item-Exposure
Control in Computerized Adaptive Testing**

Wim J. van der Linden
University of Twente, Enschede, The Netherlands

■ **Law School Admission Council**
Computerized Testing Report 02-05
May 2006

The Law School Admission Council (LSAC) is a nonprofit corporation whose members are more than 200 law schools in the United States and Canada. It was founded in 1947 to coordinate, facilitate, and enhance the law school admission process. The organization also provides programs and services related to legal education. All law schools approved by the American Bar Association (ABA) are LSAC members. Canadian law schools recognized by a provincial or territorial law society or government agency are also included in the voting membership of the Council.

© 2006 by Law School Admission Council, Inc.

All rights reserved. No part of this report may be reproduced or transmitted in any part or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission of the publisher. For information, write: Communications, Law School Admission Council, 662 Penn Street, Box 40, Newtown, PA 18940-0040.

LSAT and LSAC are registered marks of the Law School Admission Council, Inc.

This study is published and distributed by the Law School Admission Council (LSAC). The opinions and conclusions contained in these reports are those of the author and do not necessarily reflect the position or policy of the Law School Admission Council.

Table of Contents

Executive Summary	1
Abstract	1
Introduction	1
Computerized Adaptive Testing	2
Sympson-Hetter Method	2
<i>A Formal Characterization of the Sympson-Hetter Method</i>	3
<i>Conditional Item Exposure</i>	5
<i>Sampling Distribution of Item-Selection Rates</i>	5
Alternatives to the Sympson-Hetter Method	6
Simulation Studies	7
<i>First Study</i>	7
<i>Second Study</i>	9
Discussion	11
References	12

Executive Summary

In computerized adaptive testing (CAT), it is necessary to implement item-exposure control procedures to assure that the test items with the statistical characteristics that are most attractive to the item selection algorithm are not overused. The overuse of such items may compromise their security. Currently, the Simpson-Hetter (SH) method is the most popular method of item-exposure control in CAT. The method is based on a probabilistic experiment that is used to determine if an item that is selected by the CAT algorithm should actually be administered.

The implementation of the SH item-exposure control method requires that a large number of computer simulations be run in order to determine what the values of the required control parameters should be. For example, it is not uncommon to have to run 100–200 simulation studies before the values of the control parameters are determined.

In this paper, several formal properties of the SH method are presented that help us explain why the adjustment process can be slow and does not guarantee that acceptable values of the control parameters will be obtained. These properties are then used to derive several methods of item-exposure control that can serve as an alternative to the SH method. The behavior of these alternatives was studied for adaptive tests simulated from an item pool from the Law School Admission Test (LSAT). One of the alternatives showed particularly attractive behavior and resulted in acceptable values for the control parameters for all items in a fraction of the number of computer simulations required for the SH method.

Abstract

The Simpson and Hetter (SH) (1985) method is the accepted method of probabilistic item-exposure control in computerized adaptive testing. However, setting its control parameters to admissible values requires an iterative process of computer simulations that has been found to be time consuming, particularly if the parameters have to be set conditional on a realistic set of values for the examinees' ability parameter. Several formal properties of the method are given that help us explain why this iterative process can be slow and does not guarantee admissibility. In addition, some alternatives to the SH method are introduced. The behavior of these alternatives was estimated for an adaptive test from an item pool from the Law School Admission Test (LSAT). One of the alternatives showed particularly attractive behavior and converged smoothly to admissibility for all items in a relatively small number of iteration steps.

Introduction

The Simpson-Hetter (SH) (1985; see also Hetter & Simpson, 1997) method is the most popular method of item-exposure control in computerized adaptive testing (CAT). The method is based on a probabilistic experiment that is used to determine if an item that is selected should be administered. The conditional probabilities of item administration given selection of the item are the control parameters used to restrict the item-exposure rates. The values of these parameters have to be set through an iterative adjustment process in which at each step the effects of the previous adjustments are estimated using computer simulations of adaptive test administrations.

In practical settings, the use of the SH method has been found to be rather time consuming. Typically, the method is applied to control the exposure rates of the items in the pool conditional on 10–12 possible values for the examinees' ability parameters, θ . The number of iterated CAT simulations required to find admissible values for the control parameters for one θ value is generally of the same order. It is therefore not unusual to have to run some 100–150 computer simulations before the SH method can be used operationally in a CAT program. If the item pool is changed, for example, because some of its items appear to be flawed or have been the victims of a security breach, the process has to start all over again (Chang & Harris, 2002).

When using the SH method, it is regularly found that at some iteration steps, some of the exposure rates of overexposed items increase rather than decrease. Also, occasionally it appears impossible to get all exposure rates below a target value that nevertheless seems reasonable. It is unclear why we have to deal with such behavior. Because of our lack of understanding of the SH method, it is necessary to eyeball the item-exposure rates in the iterative process and use our personal judgment to decide when to stop.

This paper gives a formal characterization of the SH method. One of the key results is a constraint on the probabilities of selection of the items that applies at each step in its iterative adjustment process of the control parameters. The constraint helps us to understand why the exposure rates of the items may behave unpredictably and the SH method has difficulty finding a set of admissible values for the exposure-control parameters. We also present some alternatives to the SH method that were designed to change its less desirable features. One of these alternatives showed attractive behavior and converged directly to admissibility for all items in a relatively small number of steps.

Computerized Adaptive Testing

Let $i = 1, \dots, I$ denote the items in the pool and $k = 1, \dots, n$ denote the items in the adaptive test. In the empirical examples later in this paper, the items in the pool fit the three-parameter logistic (3PL) model. According to this model, the probability of an examinee with ability level $\theta \in (-\infty, \infty)$ on item i is equal to:

$$p_i(\theta) \equiv \Pr(U_i = 1 | \theta) \equiv c_i + (1 - c_i) \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]}, \quad (1)$$

where $b_i \in (-\infty, \infty)$, $a_i \in [0, \infty)$ and $c_i \in [0, 1]$ are parameters that represent the difficulty, discriminating power, and guessing probability on item i , respectively (Hambleton & Swaminathan, 1985).

Prior to the selection of item k , the estimator of the examinee's value for the ability parameter is updated using his/her responses on the previous $k - 1$ items. The updated estimate is denoted as $\hat{\theta}_{k-1}$. In the empirical examples below, $\hat{\theta}_{k-1}$ was the expected a posteriori (EAP) estimator.

The usual criterion to select items in CAT is the maximum-information criterion. Let $I_i(\theta)$ denote Fisher's information measure at θ ; i_k , the item in the pool that is administered as the k th item in the test; and R_k , the subset of items from the pool that is available to choose item i_k . The maximum-information criterion is given by

$$i_k \equiv \arg \max_j \{ I_{j_k}(\hat{\theta}_{k-1}) : j \in R_k \}. \quad (2)$$

For more details on the EAP estimator and maximum-information criterion as well as on alternative estimators and criteria for use in CAT, see van der Linden and Pashley (2000).

Because the criterion in Equation 2 involves maximization over a set of discrete items and the item-information functions are real, valued, and smooth, item selection in CAT typically favors the same item over small intervals of θ values. In addition, the number of intervals is usually much smaller than the size of the pool, and CAT algorithms therefore have a tendency toward overexposing a small number of items in the pool and ignoring the remaining items. The SH method has been introduced to adjust the exposure rates of items with a tendency toward overexposure.

Simpson-Hetter Method

Let S_i denote the event of selecting item i , and A_i denote the event of administering item i . For $i = 1, \dots, I$, it holds that

$$A_i \subset S_i. \quad (3)$$

Therefore,

$$P(A_i) \leq P(S_i) \quad (4)$$

and

$$P(A_i) = P(A_i, S_i) = P(A_i | S_i)P(S_i). \quad (5)$$

The goal of the SH method is to get the item-exposure rates $P(A_i)$ below an upper bound r_{\max} , that is, to get

$$P(A_i) \leq r_{\max}, \text{ for } i = 1, \dots, I \quad (6)$$

It is possible to introduce different upper bounds for different items in Equation 6, but this option will not be explored here.

The probabilities of item selection $P(S_i)$ depend on a variety of factors including: (1) the IRT model, (2) the CAT algorithm, (3) the choice of the initial item, (4) the ability estimator, (5) the composition of the item pool, and (6) the ability distribution in the population of examinees. These factors cannot be manipulated; they are fixed both by the design of the CAT and the ability distribution of the examinees in the population. However, during test administration, it is possible to manipulate the control parameters $P(A_i | S_i)$ such that the goal in Equation 6 is attained. We will refer to a set of values for the control parameters that attain the goal in Equation 6 as *admissible values*. Note that we can always reach admissibility by choosing values for $P(A_i | S_i)$ close to zero for all items. However, implicit in Equation 5 is the idea that $P(A_i)$ should be below

r_{\max} , but that for items with a tendency of overexposure, $P(A_i)$ should not be brought too far below this bound. These items generally have good measurement properties, and controlling their exposure rates too rigorously may result in substantial loss of accuracy of ability estimation. This requirement is hard to formalize but will be assumed to hold in the remainder of this paper.

The control parameters in the SH method cannot be solved analytically for a set of admissible values. Simpson and Hetter (1985) introduced an iterative rule for adjusting the values of the control parameters through a series of CAT simulations that have to be continued until admissibility is obtained. The simulations are necessary to estimate the changes in the probabilities $P(S_i)$ and $P(A_i)$ after each adjustment. The iterative rule can be defined as follows. Let t denote the iteration steps; $P^{(t)}(A_i | S_i)$, the value of the control parameter for item i at step t ; and $P^{(t)}(S_i)$ and $P^{(t)}(A_i)$, the probabilities of selecting and administering item i at step t . If the simulation at step t is completed, $P^{(t)}(S_i)$ and $P^{(t)}(A_i)$ are estimated, and for items for which the estimates of $P^{(t)}(A_i)$ do not meet Equation 6, the values of the control parameters are adjusted. The SH adjustment rule is as follows:

$$P^{(t+1)}(A_i | S_i) = \begin{cases} 1 & \text{if } P^{(t)}(S_i) \leq r_{\max} \\ r_{\max} / P^{(t)}(S_i) & \text{if } P^{(t)}(S_i) > r_{\max} \end{cases} \quad (7)$$

$i = 1, \dots, I$.

This rule is based on the following argument: If at step t item i was selected by the CAT algorithm with a probability smaller than r_{\max} , the inequality in Equation 4 implies that the item also had an exposure rate smaller than r_{\max} and does not need any control. Hence, $P^{(t+1)}(A_i | S_i)$ can be set equal to 1. However, if item i was selected with a probability larger than r_{\max} , its control parameter should have been set such that $P(A_i) = r_{\max}$. From Equation 5 it follows that this requirement would have been realized if $P^{(t)}(A_i | S_i)$ had been equal to $r_{\max} / P^{(t)}(S_i)$. Hence, $P^{(t+1)}(A_i | S_i)$ is set at this value. A more formal representation of this argument is given in Proposition 6 below.

Note that the first adjustment in Equation 7 is always upward to 1. The second adjustment is always downward to $r_{\max} / P^{(t)}(S_i)$. We will refer to upward adjustments as *positive* and downward adjustments as *negative* adjustments.

A Formal Characterization of the Simpson-Hetter Method

This section explores some of the formal properties of the SH method. The properties are presented as a set of propositions. The propositions help us to explain why the SH method shows unstable behavior in applications.

Proposition 1. For any CAT algorithm, item pool, and ability distribution, it holds that

$$\sum_{i=1}^I P(A_i) = n, \quad (8)$$

where n is the (fixed) number of items in the CAT.

This property follows from the fact that each examinee encounters n items. Because n is fixed, the average encounter rate is nI^{-1} . The equality in Equation 8 follows because the average exposure rate of the items is equal to the average encounter rate. If n is random, the right-hand side of Equation 8 should be replaced by its expectation. A random test length occurs if the CAT stops as soon as a fixed level of accuracy for the ability estimator is realized. Most CAT programs have a fixed test length though, and the remainder of this paper addresses this case.

Proposition 2. To obtain admissibility, it is necessary that

$$r_{\max} \geq nI^{-1}. \quad (9)$$

From Proposition 1, it follows that the only possible common value of $P(A_i)$ for all items is nI^{-1} . This value is a lower bound on r_{\max} because it can only be realized if the exposure-control parameters can be manipulated to have the algorithm sample items from the pool with equal marginal probabilities of selection. Real-life CAT item pools are typically 7–10 times the length of the adaptive test. However, in practice it seems impossible to get all exposure rates in the range of .10–.14, which is the corresponding range of nI^{-1} .

It is sometimes suggested that the returns on the investments in item pools for CAT should be improved by increasing the exposure rates of items that have a tendency to remain unexposed or to be hardly exposed at all. Suppose we also want to impose a lower bound r_{\min} on the exposure rates of items. The following counterpart of Proposition 2 then applies:

Proposition 3. To impose a minimum exposure rate r_{\min} on all items in the pool, it is necessary that

$$r_{\min} \leq nI^{-1} \quad (10)$$

It is impossible for a CAT algorithm to realize exposure rates equal to the upper bound in Equation 10 for all items in the item pool. Because the bound is typically a low number, Equation 10 explains why substantial numbers of items in CAT pools are never exposed. One of the alternatives to the Sympton-Hetter method proposed below is based on a minimum exposure rate larger than this upper bound, but this minimum is only imposed on items that have a tendency to be overexposed.

Proposition 4. The sum of the exposure rates remains constant across SH adjustments, that is

$$\sum_{i=1}^I P^{(t+1)}(A_i) = \sum_{i=1}^I P^{(t)}(A_i), \text{ for } t = 1, 2, \dots \quad (11)$$

This proposition is true because the equality in Proposition 1 holds within each iteration step.

The following propositions allow us to assess the possible effects of the SH adjustments in an earlier iteration step on the item exposure rates in a later step.

Proposition 5. The effects of the changes in the exposure-control parameters at Step t on the probabilities of selection of the items at Step $t + 1$ satisfy the following constraint:

$$\sum_{i=1}^I P^{(t+1)}(A_i | S_i) P^{(t+1)}(S_i) = \sum_{i=1}^I P^{(t)}(A_i | S_i) P^{(t)}(S_i). \quad (12)$$

The constraint follows from Equations 11 and 5. All factors in this constraint are known fixed constants except the probabilities $P^{(t+1)}(S_i)$. The simulation at Step $t + 1$ is conducted to estimate these probabilities. Whatever the estimates, we know that they will always have to satisfy Equation 12.

Proposition 5 explains much of the behavior of the SH method. For example, it implies that if the only adjustments of the exposure-control parameters at Step t are negative, we'll always have positive effects on the probabilities of item selection for some of the items at Step $t + 1$. Likewise, if the adjustments at Step t are positive, we'll have negative effects on some of the probabilities of selection at Step $t + 1$. Thus, more formally, if

$$P^{(t+1)}(A_i | S) < P^{(t)}(A_i | S)$$

are the only adjustments for some nonempty set $i \in U$, then

$$P^{(t+1)}(S) > P^{(t)}(S) \quad (13)$$

for some nonempty subset of items V . The same implication holds if the inequalities are reversed.

If we focus on the change in the exposure rate of one item i , it thus holds that this change is a combined result of (1) the direct effect of the change in the control parameter of item i and (2) the indirect effects of the changes of the control parameters for all other items based on the probability of selection of item i through Equation 13.

Proposition 5 also explains why the iterative process of adjustments in the SH method may show aberrant behavior. Three examples of possible aberrances are given: First, the effect of a negative (positive) adjustment of the control parameter for an item at Step t could be an increase (decrease) of the exposure rate for the *same* item at Step $t + 1$. The intended effect of the adjustment is then not observed and the SH method has not produced any progress for this item.

Second, the effect of a negative adjustment of the control parameter for one item could be the return of the exposure rate of another item to a value larger than r_{\max} . In principle, it is thus possible to have two items with exposure rates jumping back and forth between values in the intervals $[0, r_{\max}]$ and $(r_{\max}, 1]$. The SH method then has difficulty converging to admissibility.

Third, if the control parameters of several items are adjusted at the same time and some of the adjustments are negative and others positive, the effects on the probabilities of success, and thus on the

exposure rates, at the next step may largely neutralize each other. This possibility explains why the SH method sometimes progresses slowly.

The final proposition suggests a hidden assumption on which the SH method seems to rely:

Proposition 6. The adjustments in the SH method follow from the assumption that

$$P^{(t+1)}(A_i) = P^{(t+1)}(A_i | S_i)P^{(t)}(S_i), \text{ for } i = 1, \dots, I \text{ and } t = 1, 2, \dots \quad (14)$$

This proposition is immediately clear if we substitute the adjustments in Equation 7 into Equation 14 for the cases of $P^{(t)}(S_i) > r_{\max}$ and $P^{(t)}(S_i) \leq r_{\max}$. These substitutions yield the desired results, $P^{(t+1)}(A_i) = r_{\max}$ and $P^{(t+1)}(A_i) \leq r_{\max}$, respectively.

The assumption in Equation 14 would hold if the equality in Equation 5 were valid between probabilities in consecutive iteration steps. In fact, the argument directly after Equation 7 presumes the validity of Equation 5 between steps. However, Equation 5 only holds within steps. If it held between steps, it would follow from Equation 14 that the SH method always reached admissibility in one step. An alternative way to show the lack of validity of Equation 14 is to point at the fact that it would be true if $P^{(t+1)}(S_i) = P^{(t)}(S_i)$ for $i = 1, \dots, I$. However, from Equation 12 it follows that this equality only holds if $P^{(t+1)}(A_i | S_i) = P^{(t)}(A_i | S_i)$, that is, if the control parameters do not need any adjustment.

Observe that Proposition 6 contains only a sufficient condition for the SH method. It is thus possible that the method can be motivated from other assumptions. The author has not been able to find any such assumptions though.

Conditional Item Exposure

Stocking and Lewis (1998) have argued that exposure-control parameters in the SH method should be set conditional on the ability parameter θ . If they are set only marginally with respect to θ , it is still possible to have items with high exposure rates at specific ability levels because, particularly toward the end of the test, the usual item-selection criteria for CAT tend to prefer highly discriminating items at the true ability level of the examinee. Examinees may thus easily get to know some of the items before taking the test by asking examinees of comparable ability for the items they have had.

The Stocking and Lewis version of the SH method is obtained if all probabilities in Equation 7 are replaced by probabilities conditional on θ . In conditional item-exposure control, separate control parameters are to be set for a series of realistic values for θ . An advantage of conditional item-exposure control is that when setting the control parameters, no knowledge is required about the population distribution of θ . A potential disadvantage is that the control parameters have to be set conditional on true θ values but are always used conditional on estimated values. The effect of estimation error has been shown to be minor though (Stocking & Lewis, 2000).

All propositions above hold for the conditional SH method if the probabilities are defined conditional on θ . However, as indicated in the introductory section of this paper, applying the conditional method is more tedious because we need to simulate CAT conditional on a series of θ values and the number of simulations required increases linearly in the number of θ values chosen.

Sampling Distribution of Item-Selection Rates

Another factor with an impact on the iterative behavior of the SH method is the number of replications used in the CAT simulations. If the number chosen is too low, the estimates of the probabilities of selection and administration of the items have difficulty becoming stable. In particular, if the stopping criterion is based on the maximum estimated exposure rate among the items, it may be found that, due to capitalization on chance, the maximum changes in an erratic way. In fact, the author of this paper has seen the results from many applications of the SH method, typically with 250–500 CAT replications at each step, and often had the impression of watching a process with much sampling error.

It is hard to give this intuition a sound basis, though. Let $T_n = \{i_1, \dots, i_n\}$ be the set of indices of items in the pool that are administered in a CAT with n items. The event of item i being administered can be denoted by the indicator variable $I_{T_n}(i)$, which takes the value 1 if $i \in T_n$ and the value 0 otherwise. In the SH method, the simulations are conducted to estimate the probabilities $P(I_{T_n}(i) = 1)$ for all items at each iteration step. The estimates are the actual item-selection rates in the simulations. To choose a minimum number of replications, we need to know how the sampling variance of the item-selection rates depends on the number of replications. However, it is difficult to model item-selection probabilities in CAT. As noted earlier, these

probabilities depend on a variety of factors in the design of the CAT that interact in a complicated way. In fact, the variables $I_{T_i}(i)$ are closely related to the design variables in the missing-data literature on CAT (e.g., Mislavy & Chang, 2000), where it also has been found difficult to model the (conditional) probability structure of these variables.

Alternatives to the Sympon-Hetter Method

We now consider the SH method purely as an algorithm for producing admissible values for the exposure-control parameters and suggest a few alternative adjustment rules that may be more effective and do not show undesirable behavior.

The first alternative to Equation 7 is:

$$P^{(t+1)}(A_i | S_i) = \begin{cases} P^{(t)}(A_i | S_i) & \text{if } P^{(t)}(A_i) \leq r_{\max}, \\ r_{\max} / P^{(t)}(A_i) & \text{if } P^{(t)}(A_i) > r_{\max}. \end{cases} \quad (15)$$

In this adjustment rule, only negative adjustments are possible; if these adjustments lead to an exposure rate below r_{\max} , the control parameters are never adjusted back to 1. As a consequence, it is impossible to undo an earlier negative adjustment for an item by a positive adjustment at a later step. Hence, iterative processes in which alternate negative and positive adjustments lead to exposure rates that jump up and down may be avoided.

In addition, the adjustments are based on the exposure rates of the items $P^{(t)}(A_i)$ instead of their probabilities of selection $P^{(t)}(S_i)$ in two different ways: First, the size of the negative adjustment is based on $P^{(t)}(A_i)$. The inequality in Equation 4 implies

$$\frac{r_{\max}}{P(S_i)} \leq \frac{r_{\max}}{P(A_i)} \quad (16)$$

The adjustment for items with overexposure in Equation 15 is thus less rigorous than the same adjustment in the SH method. Second, from the inequality in Equation 4 it follows that the condition under which the negative adjustment in Equation 15 is applied is more restrictive than the condition in the SH method. For the latter, it is possible that items with $P^{(t)}(A_i)$ already below r_{\max} are nevertheless adjusted. Our evaluation of the adjustment rule in Equation 15 relative to the SH method is thus mixed: On the one hand, this rule may need more iteration steps because its adjustments are in smaller steps; on the other hand, fewer steps may suffice because it is better focused on items that need adjustment.

It seems, therefore, interesting to compare the adjustment rule in Equation 15 with the following rule:

$$P^{(t+1)}(A_i | S_i) = \begin{cases} P^{(t)}(A_i | S_i) & \text{if } P^{(t)}(S_i) \leq r_{\max}, \\ r_{\max} / P^{(t)}(S_i) & \text{if } P^{(t)}(S_i) > r_{\max}. \end{cases} \quad (17)$$

This rule may need fewer steps because its adjustments are more rigorous, but it is less well focused on items that need adjustment. Observe that this rule is the original SH rule modified to have only negative adjustments.

The SH rule and the two previous alternatives are based entirely either on $P(S_i)$ or $P(A_i)$. This choice gives adjusted values of the control parameters in intervals that are consistent with the conditions under which they are applied. Nevertheless, as shown in the following alternative rule, it is possible to combine more rigorous negative adjustments with application of these adjustments only to items that have exposure rates $P^{(t)}(A_i) > r_{\max}$:

$$P^{(t+1)}(A_i | S_i) = \begin{cases} P^{(t)}(A_i | S_i) & \text{if } P^{(t)}(A_i) \leq r_{\max}, \\ r_{\max} / P^{(t)}(A_i) - \delta & \text{if } P^{(t)}(A_i) > r_{\max}, \end{cases} \quad (18)$$

where $0 \leq \delta < r_{\max} / P^{(t)}(A_i)$ is a parameter set by the experimenter.

A subtle variation is possible on Equation 18 that allows us to study the effects of the presence of positive adjustments in the SH rule. Suppose we have a minimum exposure rate r_{\min} below which we do not want the exposure rates of items with a tendency of overexposure to settle. For an appropriate choice of constants δ and ε , the following adjustment rule may realize this goal:

$$P^{(t+1)}(A_i | S_i) = \begin{cases} \min\{P^{(t)}(A_i | S_i) + \varepsilon, 1\} & \text{if } P^{(t)}(A_i) < r_{\min}, \\ P^{(t)}(A_i | S_i) & \text{if } r_{\min} \leq P^{(t)}(A_i) \leq r_{\max}, \\ r_{\max} / P^{(t)}(A_i) - \delta & \text{if } P^{(t)}(A_i) > r_{\max}, \end{cases} \quad (19)$$

with $0 \leq \varepsilon < \delta < r_{\max} / P^{(t)}(A_i)$. The difference is the additional adjustment for items with $P^{(t)}(A_i) < r_{\min}$. For exposure rates below this bound, if the control parameter at the previous step was below $1 - \varepsilon$, the control parameter is adjusted positively by a quantity ε . If at the previous step the control parameter was larger, it is set at 1. The choice of ε is critical in that this parameter has to discriminate between items with a tendency to be hardly exposed, that is, items with $P^{(t)}(A_i | S_i)$ close to 1, and items that had earlier adjustments of their control parameters because of a tendency toward overexposure.

Simulation Studies

The goal of the first simulation study was to study the behavior of the alternatives to the SH method in Equation 15 and Equations 17–19 relative to the original method. The second study was suggested by the results of the first. The item pool was a previous item pool from the LSAT with 397 items fitting the IRT model in Equation 1. The adaptive test was a 30-item test from this pool. The responses on the items were simulated for examinees with θ randomly sampled from $N(0, 1)$. For efficiency, exposure control was marginal with respect to the distribution of θ . The intention was only to get an impression of the speed of convergence and the efficacy of the methods; there is no reason to expect that conditional control would lead to a different impression.

The ability of the examinee was estimated using the EAP estimator with a uniform prior on $[-5, 5]$. The initial value of the ability estimator was $\hat{\theta}_0 = 0$. The first five items were randomly selected from the 25 items with the value for the item difficulty parameters b_i closest to $\hat{\theta}_0$. These items thus had a guaranteed exposure rate equal to .25. During the simulations, the exposure-control parameters for these items were not adjusted. For each method, the adjustment process was stopped after 25 iterations. The number of examinees sampled in the CAT simulation at each step was equal to 4,000. This number was chosen to be extremely large to remove the effects of sampling error in the estimates of $P(A_i)$ and $P(S_i)$ on the behavior of the adjustment rules. The exposure-control parameters were implemented using Stocking and Lewis' (1998) multinomial experiment. For all exposure-control methods, the target for the exposure rates was $r_{\max} = .2$.

First Study

In the first study, the behavior of the SH method and the alternative adjustment rules in Equation 15 and Equations 17–19 was compared. For the rules in Equation 18 and Equation 19, we used $\delta = .15$, $r_{\min} = .10$ and $\varepsilon = .10$. The value for r_{\min} was larger than the bound in Equation 10, but, as already noted, the rule in Equation 19 imposes this value only on items that have a tendency toward overexposure.

Figures 1–3 show the number of items that violated the target value r_{\max} , the maximum exposure rate, and the average exposure rate for the violators as a function of the iteration steps. In each of these plots, the adjustment rule in Equation 18 is best. For this rule, all three performance measures were already stable at their lowest values after five iterations. The SH adjustment rule and the rule in Equation 17 produced the worst results for the number of violators, but had maximum exposure rates and average rates for the violators that tended to be second best. For the adjustment rules in Equation 15 and Equation 19, the behavior was opposite; they tended to produce numbers of violators that were second best but had the worst maximum exposure rates and average rates for the violators. Also, the behavior of these two rules was least smooth for all three performance measures. In fact, both rules showed a strong saw-tooth pattern in their maximum exposure rates and average exposure rates for the violators across the iteration steps.

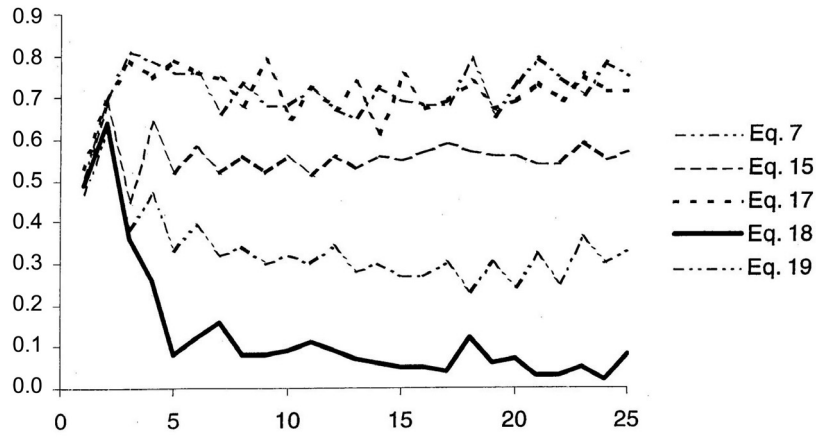


FIGURE 1. Number of violators of target exposure rate r_{max} as a function of the iteration steps for all five exposure control methods

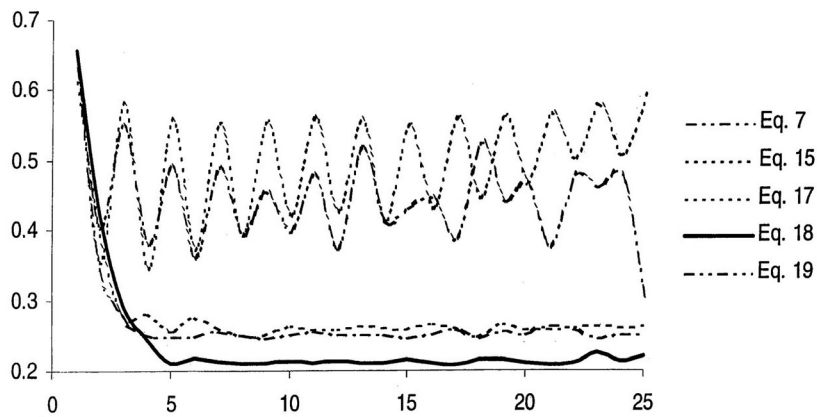


FIGURE 2. Maximum exposure rate as a function of the iteration steps for all five exposure control methods

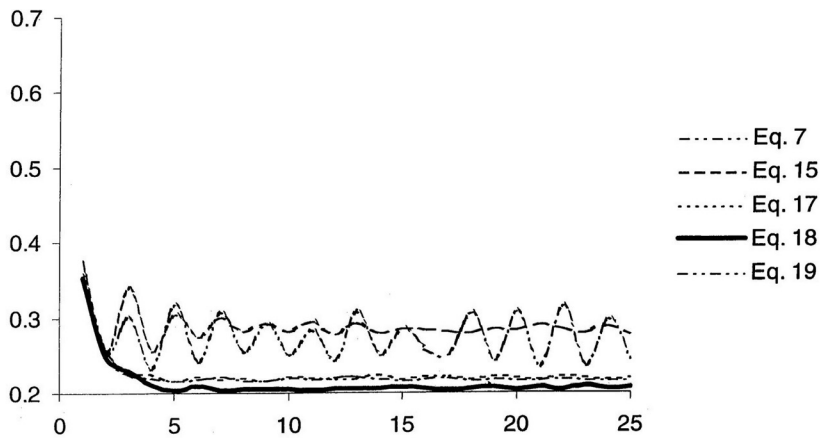


FIGURE 3. Average exposure rate of the violators as a function of the iteration steps for all five exposure control methods

Second Study

The fact that the adjustment rule in Equation 18 was uniformly best suggested a second study in which the size of the adjustment parameter δ was systematically varied. The values studied were $\delta = .05, .10, .15, .20,$ and $.25$. Because the same parameter plays a role in the adjustment rule in Equation 19, the simulations were repeated with the same parameter values for this rule.

The results for the rule in Equation 18 are given in Figures 4–6. In each of these plots, the results were identically ordered in the value of δ at nearly every iteration step, that is, the higher the value of δ , the better the results. For $\delta = .25$, the adjustment rule produced admissible exposure rates for all items at the eighth iteration step, but was already negligibly close to this result at the fifth step. This result is remarkable when compared with those for the SH method, which never produced fewer than 50 violators and had a lowest maximum exposure rate of .246 and average violation of .216 across all 25 iteration steps.

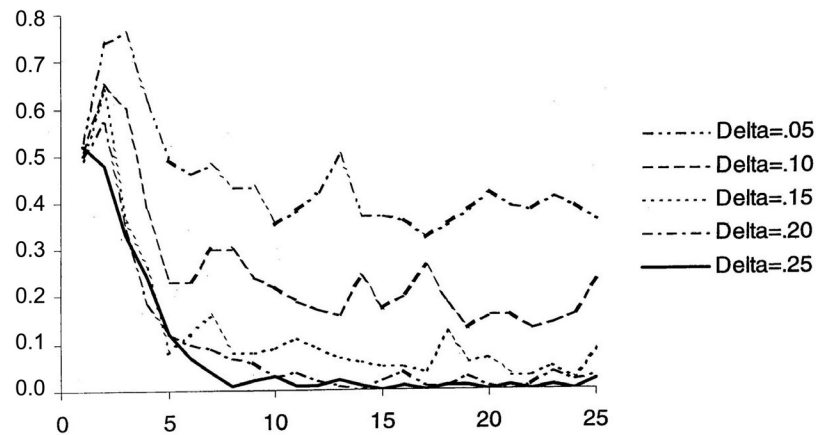


FIGURE 4. Number of violators of target exposure rate r_{max} as a function of the iteration steps for five different values of adjustment parameter δ in the exposure control method in Equation 18

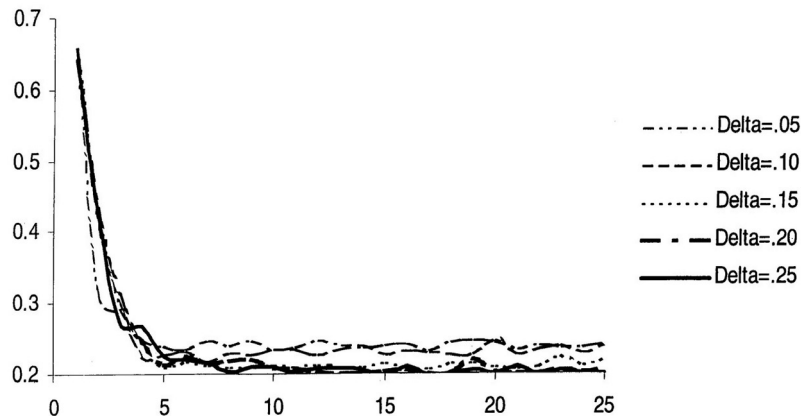


FIGURE 5. Maximum exposure rate as a function of the iteration steps for five different values of adjustment parameter δ in the exposure control method in Equation 18

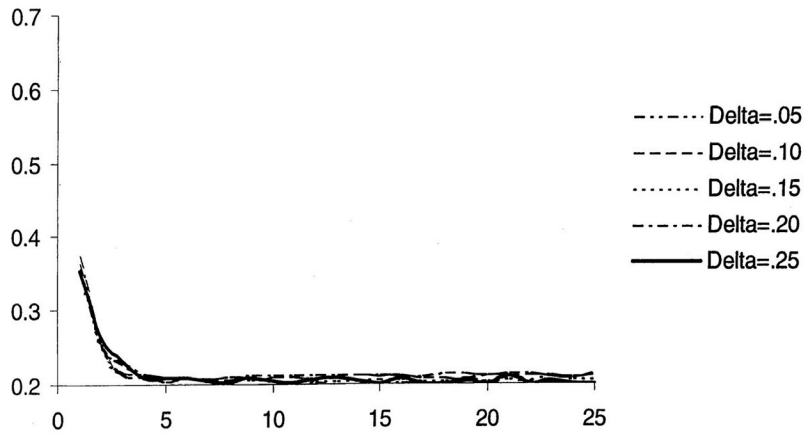


FIGURE 6. Average exposure rate of the violators as a function of the iteration steps for five different values of adjustment parameter δ in the exposure control method in Equation 18

We were not able to reproduce the above results for the same variation of δ values for the adjustment rule in Equation 19, as shown in Figures 7–9. The results for the number of violators did show the same order for increasing values of δ , but the maximum exposure rate and average rate for the violators remained as unpredictable as they were in Figures 2 and 3. The only explanation for the large difference in the results between the adjustment rules in Equations 18 and 19 is the presence of positive adjustment for items with exposure rates already admissible in the latter, that is, parameter ϵ in Equation 19.

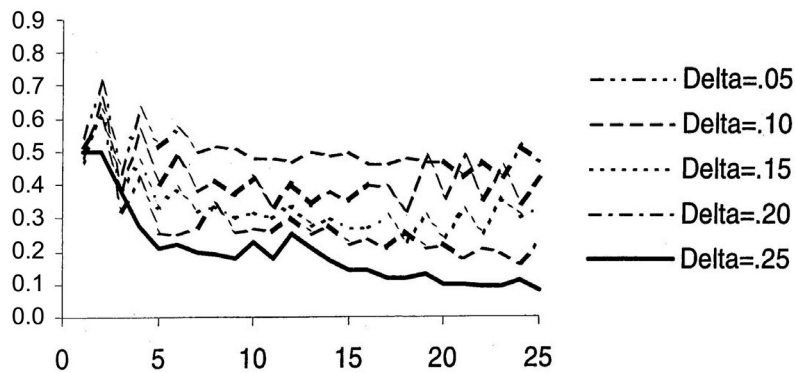


FIGURE 7. Number of violators of target exposure rate r_{max} as a function of the iteration steps for five different values of adjustment parameter δ in the exposure control method in Equation 19

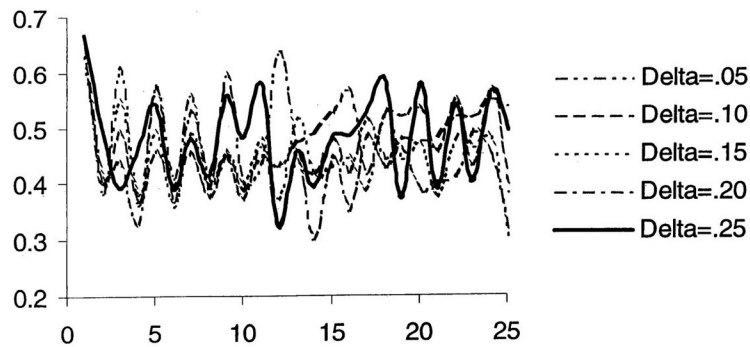


FIGURE 8. Maximum exposure rate as a function of the iteration steps for five different values of adjustment parameter δ in the exposure control method in Equation 19

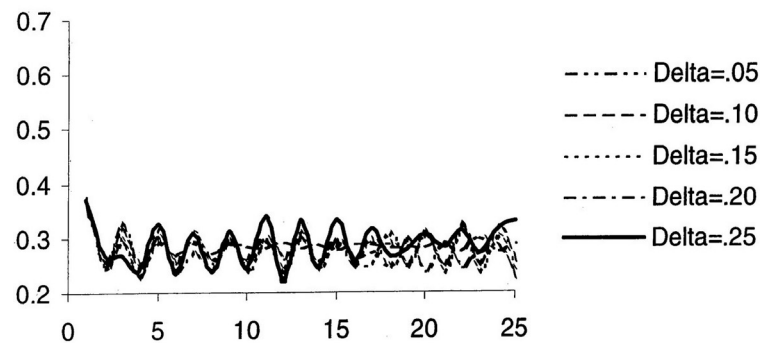


FIGURE 9. Average exposure rate of the violators as a function of the iteration steps for five different values of adjustment parameter δ in the exposure control method in Equation 19

Discussion

The formal analysis of the SH method in this paper indicated why the method may behave erratically in applications and has difficulty reaching admissibility. The simulation studies presented in this paper confirmed this behavior; the SH method did not show the fast and smooth type of convergence we would like to see. The simulation studies also showed that the adjustment rule in Equation 18 may be a practical alternative to the SH method. For adjustment parameter $\delta = .25$, this rule seems to be able to produce admissible exposure rates for all items in the pool in 5–8 iteration steps.

The combination of features of this rule that may have produced this performance is: (1) negative adjustments of the control parameters only for items with an exposure rate larger than the target value r_{\max} , (2) negative adjustments that have been made powerful because of the introduction of adjustment parameter δ , and (3) no positive adjustments for items with exposure rates already below the target value r_{\max} . The SH method misses these three features. Of course, additional studies are needed to generalize the results to other item pools and CAT algorithms. In particular, it deserves further study to find out if the presence of content constraint on item selection in CAT would force us to revise the conclusion.

Though the adjustment rule in Equation 18 seems to have the potential of a substantial reduction of the costs involved in the preparations of a new item pool for a CAT program, it is still desirable to look for less costly methods of item-exposure. An attractive alternative to the SH method seems a method based on decisions on the eligibility of the items for the examinees that are made not after the item has been selected by the algorithm but before the examinee takes the test. These decisions can be based on probability experiments that do not need any adjustment of control parameters before they can be used in an operational CAT program (van der Linden & Veldkamp, 2004).

References

- Chang, S. W., & Harris, D. J. (2002, April). *Redeveloping the exposure control parameters of CAT items when a pool is modified*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.
- Hetter, R. R., & Sympson, J. B. (1997). Item-exposure in CAT-ASVAB. In W. A. Sands, B. K. Waters & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 141–144). Washington, DC: American Psychological Association.
- Mislevy, R. J., & Chang, H. (2000). Does adaptive testing violate local independence? *Psychometrika*, *65*, 149–156.
- 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). Boston: Kluwer.
- Sympson, J. B., & Hetter, R. D. (1985, October). *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.
- van der Linden, W. J., & Pashley, P. J. (2000). Item selection and ability estimation in adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 1–25). Boston: Kluwer.
- van der Linden, W. J., & Veldkamp, B. P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics*, *29*, 273–291.