Multiple Maximum Exposure Rates in Computerized Adaptive Testing

Juan Ramón Barrada, Universidad Autonoma de Barcelona
Bernard P. Veldkamp, University of Twente
Julio Olea, Universidad Autonoma de Madrid

Computerized adaptive testing is subject to security problems, as the item bank content remains operative over long periods and administration time is flexible for examinees. Spreading the content of a part of the item bank could lead to an overestimation of the examinees' trait level. The most common way of reducing this risk is to impose a maximum exposure rate (r^{max}) that no item should exceed. Several methods have been proposed with this aim. All of these methods establish a single value of r^{max} throughout the test. This study presents a new method, the multiple-r^{max} method, that defines as many values of r^{max} as the number of items presented in the test. In this way, it is possible to impose a high degree of randomness in item selection at the beginning of the test, leaving the administration of items with the best psychometric properties to the moment when the trait level estimation is most accurate. The implementation of the multiple-r^{max} method is described and is tested in simulated item banks and in an operative bank. Compared with a single maximum exposure method, the new method has a more balanced usage of the item bank and delays the possible distortion of trait estimation due to security problems, with either no or only slight decrements of measurement accuracy.
Index terms: computerized adaptive testing, item exposure control, test security, item selection

Computerized adaptive testing (CAT) of knowledge, abilities, and skills offers several advantages. CATs are administered individually and they are flexible. Moreover, they are more efficient than traditional paper-and-pencil testing in that the difficulty of the items can be adapted to the proficiency of the candidate (Segall, 2004).

However, CATs have also been criticized. First of all, they are subject to security problems. When they are online, they are vulnerable to (organized) item theft. Candidates might memorize items and publish them on the Internet or simply share them with friends who might take the CAT in the future (Chang, 2004). A second problem is related to item bank usage. Many items in the banks are rarely selected for administration, because most item selection rules favor other items for their better measurement qualities. Thus, both time and money are wastefully invested in developing them.

Both problems can be formulated in terms of exposure of individual items: security problems are related to variance in the exposure rates of the items (Chen, Ankenmann, & Spray, 2003); poor item bank usage is related to an underexposure of less popular items. To deal with these problems, various exposure control methods have been proposed, the most popular being that of Sympson and Hetter (1985). Numerous modifications of this method have been presented (Stocking & Lewis, 1998; van der Linden, 2003). Chang and Ying (1999) proposed the alpha-stratified method;

Applied Psychological Measurement, Vol. 33 No. 1, January 2009, 58–73

DOI: 10.1177/0146621608315329

Revuelta and Ponsoda (1998) the progressive method, which focuses on underexposure problems; and more recently, van der Linden and Veldkamp (2004, 2007) developed the item-eligibility method.

In all these methods, the following trade-off can be found: the greater the emphasis on exposure control, the greater are the costs in terms of measurement precision (Way, 1998). From the inverse point of view, the more accurate the CAT, the higher are the risks to the item bank. In fact, this study deals with a multiple-criteria decision-making problem. The first criterion is measurement precision; the second, exposure control. Therefore, the challenge in developing or selecting exposure control methods lies in finding the method that performs best with respect to both measurement precision and observed exposure rates or test overlap.

In this article, a new method for exposure control, the multiple-$r^{\max}$ method (MRM), is described. In this method, exposure control parameters are varied throughout the test administration. It is argued that increased item bank usage can be achieved with this method, with either no or at the most only minimal increments in measurement errors.

The common method for improving bank security is to control the maximum exposure rate. First, one of the methods for doing so, the item-eligibility method (van der Linden & Veldkamp, 2004), is described. After that, some of its limitations are shown and then the rationale of the MRM method and its implementation are presented. Two simulation studies are described, one with randomly generated item banks and the other with an operative item bank.

Item-Eligibility Method

The goal of methods that control the *maximum exposure rate* is to set all *item exposure rates* below a maximum exposure rate, $r^{\max}$, fixed beforehand by the testing agency:

$$P(A_i) \leq r^{\max}, \tag{1}$$

where $Q/n \leq r^{\max} \leq 1$ (Chen et al., 2003), $P(A_i)$ is the probability of administering the ith item, Q is the number of items to be administered, and n is the item bank size. All the methods introduce exposure control parameters for the items. The first method presented was the Sympson−Hetter method (Sympson & Hetter, 1985). This approach involves a time-consuming process to fix the exposure control parameters (Barrada, Olea, & Ponsoda, 2007; Chen & Doong, 2008; van der Linden, 2003). Some methods have been proposed that adapt the control parameters on the fly. The restricted method (Revuelta & Ponsoda, 1998) has this characteristic, but some drawbacks of the method have been described (Chen, Lei, & Liao, in press). Recently, van der Linden and Veldkamp (2004, 2007) described the item-eligibility method, which the present study uses as a benchmark for comparison with the MRM method.

In the item-eligibility method, two events are defined: (a) item i is eligible for the examinee (E_i) and (b) item i is administered (A_i). Exposure control is achieved by restricting the proportion of examinees for which an item can be eligible. This proportion is $P(E_i)$.

The $P(E_i)$ values are adapted on the fly for each new examinee. The parameters for the $(j+1)$-th examinee—$P(E_i^{j+1})$—are calculated using the following equation:

$$P(E_i^{j+1}) = \begin{cases} 1 & \text{if} \quad P\left(A_i^{1..j}\right)/P\left(E_i^j\right) \leq r^{\max} \\ r^{\max} P\left(E_i^j\right)/P\left(A_i^{1..j}\right) & \text{if} \quad P\left(A_i^{1..j}\right)/P\left(E_i^j\right) > r^{\max} \end{cases}, \tag{2}$$

where $P(A_i^{1..j})$ is the exposure rate (probability of administration) of the ith item in the range of examinees between the first and the jth examinee.

For each examinee, a subset of eligible items is formed before any item is administered. For each item, a random number belonging to the uniform interval (0, 1) is generated. Only if that number is smaller than $P(E_i^{j+1})$ is that item eligible. During the administration of the test, only eligible items can be administered. This way, except for a few random exceptions, all items have exposure rates equal to or below $r^{\max}$.

The methods presented to date for controlling the maximum exposure rate share several drawbacks. Assuming that the item with maximum Fisher information for the estimated trait level is selected and items calibrated according to the three-parameter logistic model are used, items with a high *a* parameter and a low *c* parameter from the beginning of the test will be chosen (Barrada, Olea, Ponsoda, & Abad, 2006) when the estimation of the trait level is unstable and measurement error is high. As items in the bank with this combination of parameters are infrequent, the quality of items measured by the information they provide will decline as the test goes on (Revuelta & Ponsoda, 1998).

Even with $r^{\max}$ values close to the minimum possible, the mean of the *a* parameter of the items administered is still above the mean of the *a* parameter in the bank (Barrada et al., 2007). The methods of restriction of the maximum exposure rate reduce the exposure rate of overexposed items while increasing the exposure rate of items whose exposure rates are smaller and closer to $r^{\max}$.

Hau and Chang (2001) have shown that it is advisable to increase the value of the *a* parameter of the items administered as the test goes on, instead of reducing it. In fact, the random selection of items at the beginning of the test and subsequent selection based on Fisher information means no reduction or a very small reduction (Barrada, Olea, Ponsoda, & Abad, in press; Li & Schafer, 2005; Revuelta & Ponsoda, 1998) in measurement accuracy.

One method for balancing the item exposure rates would involve establishing as many maximum exposure rates as items to be administered (Q). A proposal for doing so is presented in the next section.

A New Method: The Multiple-$r^{\max}$ Method

The goal of the multiple-$r^{\max}$ method (MRM) can be seen in equation (3):

$$P(A_{i,1..q}) \le r^{\max}_{1..q}, \tag{3}$$

where $r^{\max}_{1..q}$ is the desired maximum exposure rate until the qth item and $P(A_{i,1..q})$ is the exposure rate of the i item considering the first q items in the test.

The definition of the $r^{\max}_{1..q}$ values is subject to the following restrictions:

$$r^{\max}_{1..q+1} > r^{\max}_{1..q}, \tag{4}$$

$$r^{\max}_{1..q} \ge q/n, \tag{5}$$

and

$$r^{\max}_{1..Q} \le 1. \tag{6}$$

If $r^{\max}_{1..q+1}$ was allowed to be equal to $r^{\max}_{1..q}$, those items with $P(A_{i,1..q})$ equal to $r^{\max}_{1..q}$ could not be administered in the $(q+1)$-th position of the test. This is avoided by the first restriction. The other two restrictions mark the limits between which the $r^{\max}_{1..q}$ values have to be.

The lowest $r^{\max}_{1..q}$ is imposed at the beginning of the test. In this way, it is possible to avoid selecting all items with high *a* and low *c* parameters when estimation of trait levels is still unstable. The values of $r^{\max}_{1..q}$ increase during CAT administration, which implies that more items become eligible.

The MRM model presented by inequalities 3 to 6 is not a stand-alone method of item exposure control but a general structure that needs to be combined with a method for controlling maximum exposure rates. The next section explains how the MRM has been implemented, combined with the item-eligibility method (van der Linden & Veldkamp, 2004).

Implementation of MRM

Some examples are first presented to illustrate the implementation of this new method. Imagine the following condition: $r^{\max}_{1..q-1}$ equals 0.24 and $r^{\max}_{1..q}$ equals 0.25. Which maximum exposure rate can therefore be tolerated for the qth position? Not 0.25. If item 1 is exposed to 24% of the examinees until the $(q-1)$-th position and exposed to 25% in the next position, the total exposure rate of that item would clearly be over our limit.

The two maximum exposure rates for an item position need to be differentiated. First, the maximum exposure rate acceptable during the first qth items is considered. As stated above, this first maximum exposure rate is termed $r^{\max}_{1..q}$. Second, the maximum exposure rate tolerable in the qth position.

This tolerable exposure rate could be defined as the difference between $r^{\max}_{1..q}$ and $r^{\max}_{1..q-1}$. In this case, in the present example, the tolerable rate for the qth item position would be 0.01. Consider item 2, exposed to 5% of the examinees until the $(q-1)$-th position, not reaching $r^{\max}_{1..q}$. If an exposure rate of only 0.01 is tolerated in the qth position, the total exposure rate of this item would be markedly below the limit, losing tolerable usage of that item. This article attempts to satisfy equation (3) without overrestricting exposure rates. It can be seen with the example of item 2 that the tolerable rate should be dependent on the exposure rate in the previous item position.

One option might be to calculate the tolerable exposure rate for the qth position equal to $r^{\max}_{1..q}$ minus the actual exposure rate until the $(q-1)$-th position. For item 1, this value is equal to 0.01 and for item 2 it is equal to 0.2. For these two items there is no problem with this definition of the tolerable exposure rate for the qth position. But imagine items 3 and 4. Item 3 is exposed to 0.245 of the examinees until the $(q-1)$-position and item 4 to 0.26. Both exposure rates are greater than $r^{\max}_{1..q-1}$ and in the case of item 4, greater than $r^{\max}_{1..q}$. Observed exposure rates higher than those desired are possible because the process includes a random component. If the definition of tolerable exposure rate as $r^{\max}_{1..q}$ minus the actual exposure rate until the $(q-1)$-th position is applied, the figures for items 3 and 4 would be 0.005 and -0.01. It is meaningless to set a negative value for the tolerable exposure rate in an item position. One option could be to fix negative values at zero, but the convenience of not doing so has been defended in this article (equation (4)). Both the MRM method and the item-eligibility method adapt all the parameters on the fly to try to satisfy the restrictions imposed, so it makes sense to suppose that when a new examinee is tested, the observed exposure rates of the items that exceed exposure limits will fall to the limits fixed. Considering that this control is achieved, the tolerable item exposure rate would then be 0.01 for items 3 and 4. In this way, it can be seen how the tolerable exposure rate will depend on the estimation of the exposure rate in the previous item position when a new examinee is tested. This estimation is made as in equation (7):

$$\hat{P}\left(A^{1..j+1}_{i,1..q}\right) = \begin{cases} P\left(A^{1..j}_{i,1..q}\right) & \text{if} \quad P\left(A^{1..j}_{i,1..q}\right) < r^{\max}_{1..q} \\ r^{\max}_{1..q} & \text{if} \quad p\left(A^{1..j}_{i,1..q}\right) \geq r^{\max}_{1..q} \end{cases} \tag{7}$$

Table 1
Examples of Definition of Maximum Exposure Rate in a Position of the Test

	a		*b*		*c*		*d*	
q	1	2	1	2	1	2	1	2
$r_{1..q}^{\max}$	0.05	0.15	0.05	0.15	0.05	0.15	0.05	0.15
$P(A_{i,1..q}^{1..j})$	0.03	0.15	0.05	0.20	0.10	0.15	0.18	0.20
$t_{i,q}^{j+1}$	0.05	0.12	0.05	0.10	0.05	0.10	0.05	0.10

As can be seen, it is assumed that in the event the exposure rate for one examinee exceeds the maximum exposure rate, the exposure control method will be able to restrict the exposure rate to $r_{1..q}^{\max}$ for the next examinee.

The tolerable rate in the qth position in the test for the ith item in the bank for the $(j+1)$-th examinee is referred to as $t_{i,q}^{j+1}$. The value of $t_{i,q}^{j+1}$ will be $r_{1..q}^{\max}$ minus the estimation of $P(A_{i,1..q}^{1..j+1})$. Thus, the equation for calculating $t_{i,q}^{j+1}$ is as follows:

$$t_{i,q}^{j+1} = r_{1..q}^{\max} - \hat{P}\left(A_{i,1..q-1}^{1..j+1}\right) = r_{1..q}^{\max} - \min\left[r_{1..q-1}^{\max}, P\left(A_{i,1..q-1}^{1..j}\right)\right]. \tag{8}$$

Table 1 shows four examples presenting how $t_{i,q}^{j+1}$ is calculated. In example d, it is clear why it is better to use equation (7) for calculating $\hat{P}(A_{i,1..q}^{1..j+1})$ instead of simply making $\hat{P}(A_{i,1..q}^{1..j+1})$ equal to $P(A_{i,1..q}^{1..j})$. With the option chosen, it is impossible for $t_{i,q}^{j+1}$ to be negative or equal to zero.

When MRM is combined with the item-eligibility method, the control parameters for item i for the $(j+1)$-th examinee and the qth item position is calculated according to the following equation:

$$P\left(E_{i,q}^{j+1}\right) = \begin{cases} 1 & \text{if} \quad P\left(A_{i,1..q}^{1..j}\right)/P\left(E_{i,q}^{j}\right) \le t_{i,q}^{j+1} \\ P\left(E_{i,q}^{j}\right)t_{i,q}^{j+1}/P\left(A_{i,1..q}^{1..j}\right) & \text{if} \quad P\left(A_{i,1..q}^{1..j}\right)/P\left(E_{i,q}^{j}\right) > t_{i,q}^{j+1} \end{cases} \tag{9}$$

As can be seen, this equation is similar to the one used for calculating the $P(E_i)$ parameters in the item-eligibility method but $r^{\max}$ is replaced with $t_{i,q}^{j+1}$.

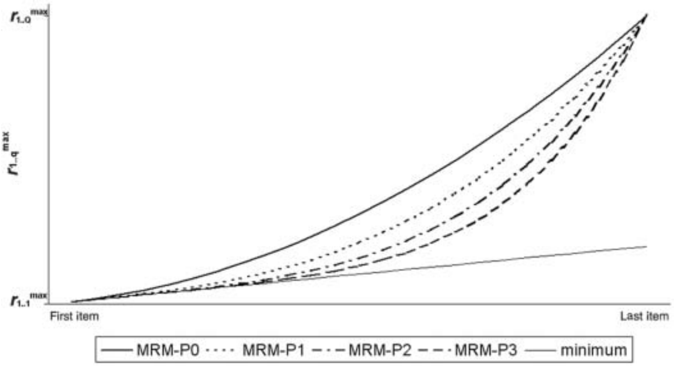
Once the control parameters are calculated, it is possible to define which part of the bank is eligible for each item position. This is done as explained in the item-eligibility method, with the difference that eligibility is not defined for the test but for each item position. For doing so, n random numbers in the uniform interval (0, 1) are generated, one for each item position, and if these numbers are lower than the control parameters the items are eligible for those item positions.

If content constraints are incorporated into the test (van der Linden, 2005), there is a possibility of no feasible test existing. This would happen when the eligible items cannot meet the content specifications. Van der Linden and Veldkamp (2004) discuss how to incorporate the probability of infeasibility into the item-eligibility method. For an item bank correctly constructed, this probability is considered to be very small and this element is not introduced in the present method.

A Possible Function for Defining the Values of $r_{1..q}^{\max}$

The random selection of items at the beginning of the test has a small impact on measurement accuracy (Barrada et al., in press; Li & Schafer, 2005; Revuelta & Ponsoda, 1998). Thus, it seems

Figure 1
Examples of Functions Relating Item Position to $\boldsymbol{r_{1..q}^{\max}}$ for Four Acceleration Parameters and the Minimum Admissible Values of $\boldsymbol{r_{1..q}^{\max}}$



appropriate to strongly adjust the $r_{1..q}^{\max}$ values for the first items to the minimum admissible values (equation (5)), as this would improve the balanced usage of the item bank. The value for $r_{1..Q}^{\max}$ would be set to the value that for security reasons the testing agency considers suitable. The most logical option is to fix $r_{1..Q}^{\max}$ equal to the value that $r^{\max}$ would have if a method with a single $r^{\max}$ was applied.

A possible function for defining the $r_{1..q}^{\max}$ values that makes $r_{1..1}^{\max}$ equal to n^{-1} and leaves the freedom to set $r_{1..Q}^{\max}$ is as follows:

$$r_{1..q}^{\max} = \begin{cases} \frac{q}{n} & \text{if} \quad q = 1 \\ \left[1 + \frac{\left(r_{1..Q}^{\max} \Big/ Q/n - 1 \right) \sum_{h=2}^{q} (h-1)^s}{\sum_{h=2}^{Q} (h-1)^s} \right] \frac{q}{n} & \text{if} \quad q \neq 1 \end{cases}, \tag{10}$$

where h is a dummy variable only used for calculations and s is the acceleration parameter defining the speed with which $r_{1..q}^{\max}$ separates from the minimum possible values for approaching $r_{1..Q}^{\max}$. Examples of this function are shown in Figure 1.

With an acceleration parameter equal to zero, the ratio between $r_{1..q}^{\max}$ and q/n (the minimum admissible value; equation (5)) increases in a linear fashion from 1 to $r_{1..Q}^{\max} \Big/ Q/n$. The higher the value of the s parameter, the lower the speed with which the $r_{1..q}^{\max}$ values increase.

It is important to note that obtaining a homogeneous distribution of the exposure rates by adjusting the values of $r_{1..q}^{\max}$ to their minimum admissible values is not equivalent to selecting items randomly. By setting $r_{1..q}^{\max}$ equal to q/n, the exposure rates for the overall population are homogenized, although when considering the exposure rates conditional on trait levels there will be a variance in the distribution of exposure rates that would not occur with random selection. Owing to this, the measurement error achieved with methods of restriction of maximum exposure rate—even though the maximum exposure rates are fixed at the minimum possible—will be lower than that found with random selection.

Two simulation studies were carried out to evaluate the performance of MRM as compared with a method using a single value of $r^{\max}$, the item-eligibility method. In the first study, randomly generated item banks were used, whereas in the second a currently operative item bank for the assessment of knowledge of English grammar (Olea, Abad, Ponsoda, & Ximénez, 2004) was used.

Simulation Study 1

Method

Item banks and test length. Ten item banks of 500 items were generated. The distributions for the parameters were as follows: for a, $N(1.2, 0.25)$; for b, $N(0, 1)$; and for c, $N(0.25, 0.02)$. The test length was fixed at 25 items.

Trait level of the examinees. This study aimed to obtain the results for the overall population and conditional on several θ values. It was decided to sample nine θ values, ranging from -2 to 2 in steps of 0.5. To do so, a pool of examinees was constructed with the following two conditions: (a) the number of examinees in each θ value had to be proportional to the density at that point assuming a distribution $N(0, 1)$; (b) the minimum number of examinees for any θ value had to be equal to 1,000, considering this number large enough to obtain stable results. In this way, the pool of examinees was composed of 36,193 elements: for θ equal to -2 and 2, 1,000 examinees were sampled; for θ equal to -1.5 and 1.5, 2,399 examinees; for θ equal to -1 and 1, 4,482 examinees; for θ equal to -0.5 and 0.5, 6,521 examinees; and for θ equal to 0, 7,389 examinees. The trait level of each examinee simulated was randomly extracted without replacement from this pool.

Estimation of trait level and item selection rule. The starting $\hat{\theta}$ was fixed at 0. The estimator of θ was the expected a posteriori (EAP; Bock & Mislevy, 1982) estimator with a uniform prior over $[-4, 4]$. The selection algorithm most widely used in CATs was used: selecting the item with maximum Fisher information at the current estimated trait level.

$r_{1..q}^{\max}$ *values.* These values were adjusted for each item position in the test as shown in equation (10). The $r_{1..Q}^{\max}$ value was set equal to 0.25. Four values were used for the acceleration parameter: 0, 1, 2, and 3. In the item-eligibility method, $r^{\max}$ was equal to 0.25.

Performance measures. Six variables were used for the comparison between methods: (a) observed maximum exposure rates; (b) exposure rates of the items at the end of the test; (c) overlap rate, as defined in equation (11) (Chen et al., 2003); (d) RMSE (root mean square error), as shown in equation (12); (e) bias, calculated following equation (13); and (f) the information provided by the items for the real trait level of the examinee. RMSE and bias were calculated both for the whole set of simulees and conditional on the different θ values.

The overlap rate was

$$\hat{T} = \frac{n}{Q} S_{P(A)}^2 + \frac{Q}{n}, \tag{11}$$

where $\hat{T}$ is the large-sample approximation of the overlap rate (Chen et al., 2003) and $S_{P(A)}^2$ is the variance of the item exposure rates.

The RMSE and bias were

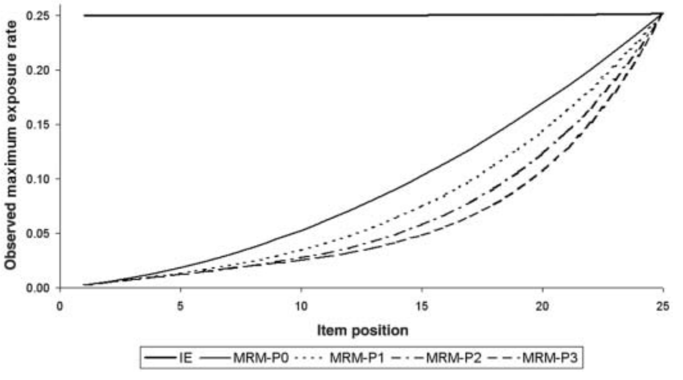
$$\text{RMSE} = \left(\sum_{g=1}^{m} \left(\hat{\theta}_g - \theta_g \right)^2 \Big/ m \right)^{1/2} \tag{12}$$

and

$$\text{Bias} = \sum_{g=1}^{m} \left(\hat{\theta}_g - \theta_g \right) \Big/ m, \tag{13}$$

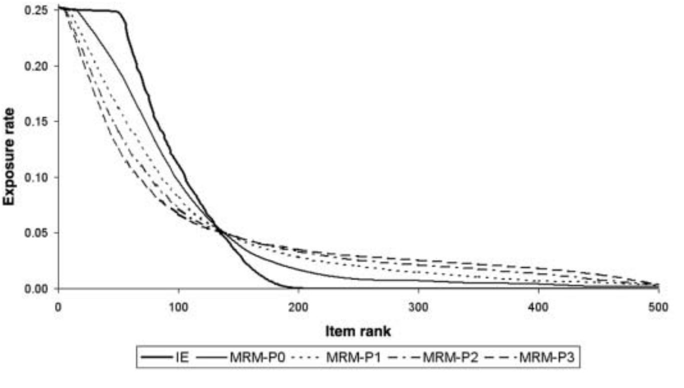
where m is the number of examinees, $\hat{\theta}_g$ is the estimated trait level for the gth examinee, and θ_g is the real trait level.

Figure 2
Observed Maximum Exposure Rate According to Item Position for the Theoretical Item Banks



Note: IE = item-eligibility method.

Figure 3
Exposure Rates of Items for Theoretical Item Banks



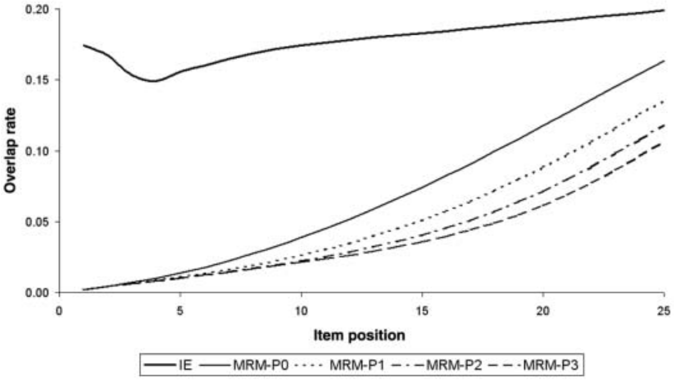
Note: The items are ordered according to their exposure rates. IE = item-eligibility method.

Results

Figure 2 shows the maximum exposure rates according to the item position in the test. With the item-eligibility method, the maximum exposure rate is already r^{max} for the first item and remains constant through the test. The MRM method shows the desired pattern for the different acceleration parameters studied: maximum rate at the beginning of the test is very low and increases as the number of items administered increases. The magnitude of this increase is controlled by the acceleration parameter. Both the item-eligibility method and the MRM method succeed in controlling the maximum exposure rate at the desired level.

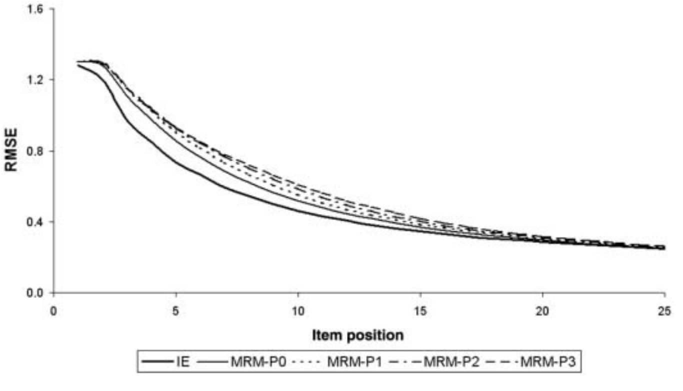
As expected, the MRM method leads to a more homogeneous distribution of exposure rates, as can be seen in Figure 3. Although with the item-eligibility method about 60% of the items in the bank are never used, with the MRM method no item has an exposure rate equal to zero. MRM also

Figure 4
Overlap Rate According to Item Position for the Theoretical Item Banks



Note: IE = item-eligibility method.

Figure 5
RMSE According to Item Position for the Theoretical Item Banks



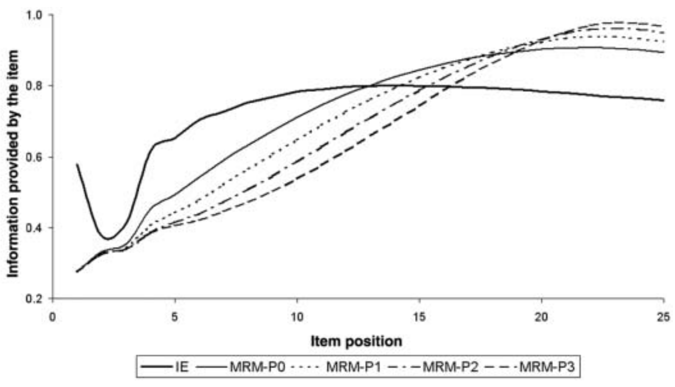
Note: IE = item-eligibility method; RMSE = root mean square error.

reduces the proportion of items with rates close to the maximum limit established. The greater the value of the acceleration parameter, the greater are the improvements in relation to under- and overexposure.

Greater exposure control means a reduction in the overlap rate achieved with the MRM method, which is always lower than the overlap rate obtained with the item-eligibility method, as can be seen in Figure 4. The overlap correlates negatively with the acceleration parameter. The differences between the item-eligibility method and MRM are greater at the beginning of the test and decrease as more items are administered.

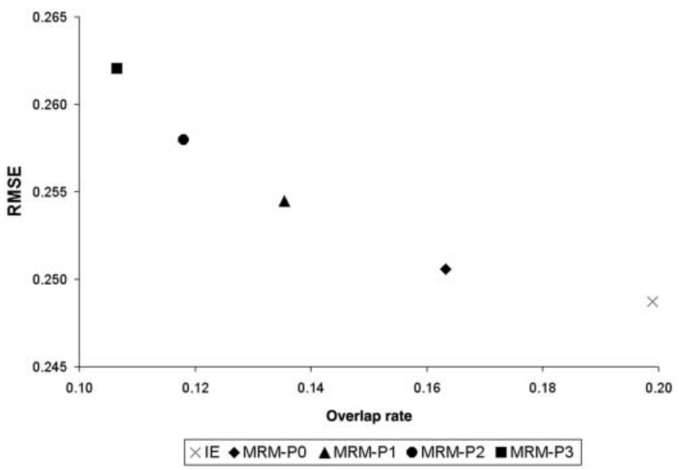
The lower exposure rate of MRM at the beginning of the test can be an additional advantage of the method. The distortion in the estimated trait level will be greater if the items that the examinee has previous knowledge of are at the beginning of the test, and especially if their *a* parameter is high (Chang & Ying, in press), as occurs with the item-eligibility method.

Figure 6
Fisher Information Provided by the Item According to Item Position for the Theoretical Item Banks



Note: IE = item-eligibility method.

Figure 7
Overlap Rate and RMSE for the Theoretical Item Banks

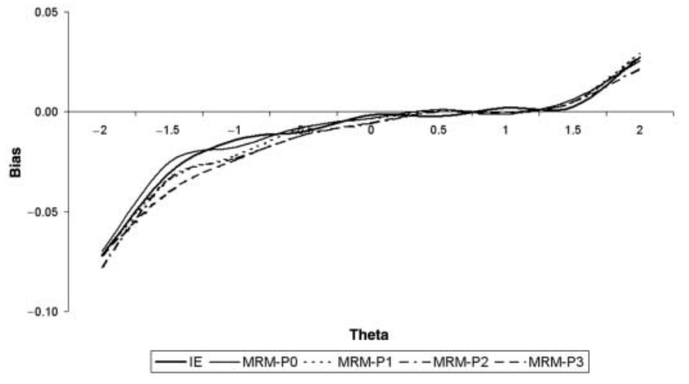


Note: IE = item-eligibility method; RMSE = root mean square error.

The effect of this greater exposure control on accuracy can be seen in Figure 5. At the beginning of the test, the MRM method offers a higher RMSE than the item-eligibility method, but this difference between them quickly falls, and by the end of the test it is almost unnoticeable.

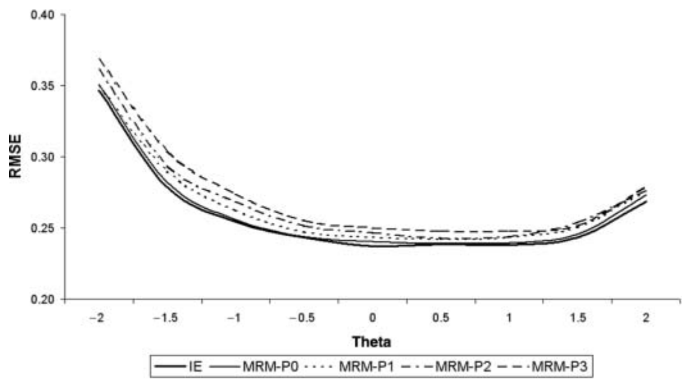
The reason why it is possible to improve the bank security with almost no impact on accuracy can be seen in Figure 6, which shows the plotting of the mean information provided by each item for the examinee's real trait level. In the item-eligibility method, the information provided by each item increases in the first half of the test as the estimation approaches the real trait level. The exception of the second item in the test, when the information provided is reduced, is because the

Figure 8
Bias According to **θ** Values for the Theoretical Item Banks



Note: IE = item-eligibility method.

Figure 9
RMSE According to **θ** Values for the Theoretical Item Banks

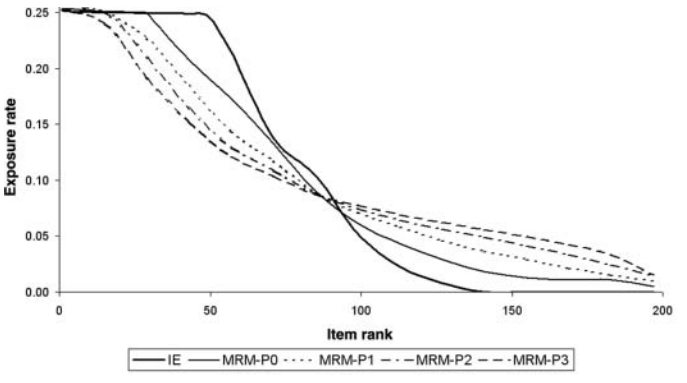


Note: IE = item-eligibility method; RMSE = root mean square error.

estimation after administration of just one item is necessarily far removed from the mean of the trait level in the population. For the second half of the test, the information provided by each item in the item-eligibility method reduces with each new item presented, as the trait level estimated is more stable and the highly informative items have already been used. Considering the first items of the test, the items presented with the MRM method are less informative than the items administered with the item-eligibility method. However, as far as the latter part of the test is concerned, there are still high-quality items available in the bank. This means that the information gathered with the two methods is similar and explains the small difference in RMSE.

Because the relevant point in practical settings is what is obtained at the end of the test, Figure 7 shows the overlap rates and RMSE for the two methods after 25 items. With regard to overlap, it shows how the MRM method clearly improves item bank security, compared with the item-eligibility method. This improvement increases as the acceleration parameter increases. With regard

Figure 10
Exposure Rates of Items for the Operative Bank



Note: The items are ordered according to their exposure rates. IE = item-eligibility method.

to RMSE it is clear how, as noted above, the differences are small and never greater than 0.015. The greater the value of the acceleration parameter, the greater is the RMSE.

Figure 8 shows the bias for the different methods according to the θ values sampled. The main differences are located for low trait-level values and can be considered as small. The higher the acceleration parameter, the higher is the bias. For θ values greater than 0, no differences are distinguishable. The same can be said for differences between the item-eligibility method and the MRM method with an acceleration parameter equal to 0.

The RMSE for the item-eligibility method is lower than the RMSE for the MRM method when acceleration parameters more than 0 are considered, as can be seen in Figure 9. When MRM with an acceleration parameter of 0 is compared with the item-eligibility method, the differences are negligible. The RMSE correlates positively with the acceleration parameter.

In short, the MRM method, compared with the item-eligibility method, allows for greater exposure control. It reduces the overlap rate and the number of items with exposure rates close to the limit rate and increases the exposure rate of the underexposed items. Moreover, these advantages are achieved with very little impact on measurement accuracy. The explanation is that the MRM method saves the most informative items for the latter part of the test, when the trait estimation is more accurate.

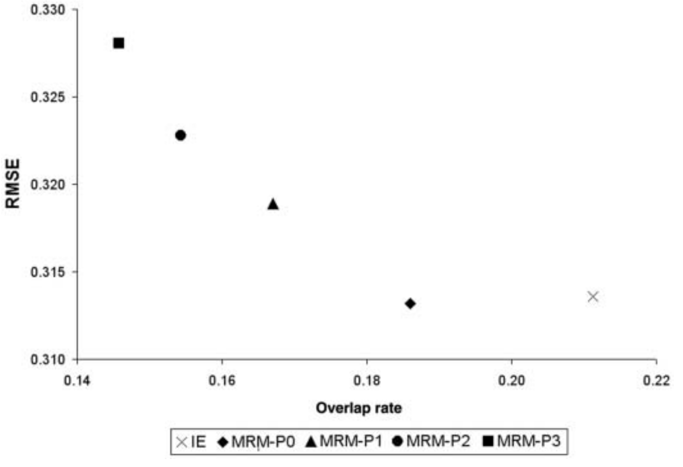
It seems clear that in the case of randomly generated item banks, the MRM method is an option that improves the security control of the item bank when compared with the methods that work with just one maximum exposure rate. To test whether the results of this study can be generalized, MRM was applied to a currently operative item bank.

Simulation Study 2

Method

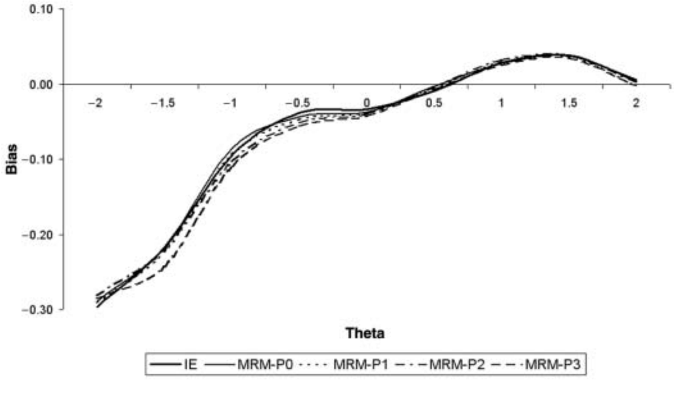
The method of this second study is similar to that of the first except in certain aspects. It used an item bank for assessing knowledge of English grammar, eCAT (Olea et al., 2004), used in human resources contexts for personnel selection and promotion. The bank has 197 items. This small size means that security issues are especially relevant for eCAT. The mean, standard deviation,

Figure 11
Overlap Rate and RMSE for the Operative Bank



Note: IE = item-eligibility method; RMSE = root mean square error.

Figure 12
Bias According to **θ** Values for the Operative Bank



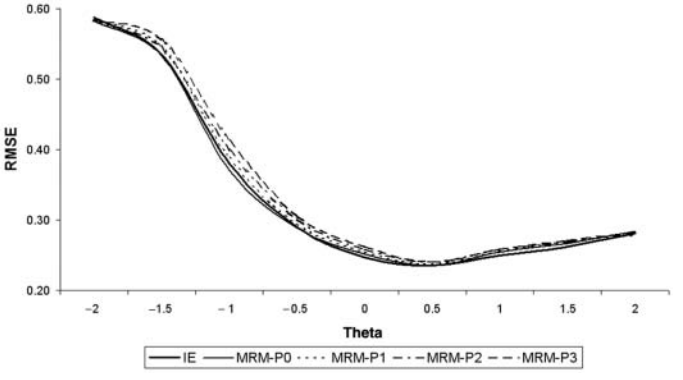
Note: IE = item-eligibility method.

maximum, and minimum for the *a*, *b*, and *c* parameters were (1.3, 0.32, 2.2, 0.43), (0.23, 1, 3.42, −2.71), and (0.21, 0.03, 0.29, 0.11), respectively. Test length was fixed at 20 items. Although in practice the test length depends on the needs of the companies that use it, this length is the one usually chosen. The maximum exposure rate was fixed at 0.25, as in the first study.

Results

Only the results for the exposure rates, overlap rates, RMSE, and bias at the end of the test are shown as these are the relevant data in a practical context.

Figure 13
RMSE According to **θ** Values for the Operative Bank



Note: IE = item-eligibility method; RMSE = root mean square error.

Figure 10 shows the exposure rate for the different methods. Basically, the results of Study 1 are replicated. Although with the item-eligibility method a considerable part of the item bank is never used, with the MRM method there is no item with an exposure rate equal to zero. Also, the MRM method reduces the proportion of items with an exposure rate close to the limit rate. These effects are more marked as the value of the acceleration parameter is increased.

Figure 11 shows the overlap rate and RMSE values. In accordance with that seen with the exposure rates, the item-eligibility method has the highest overlap rate. The greater the value of the acceleration parameter, the lower is the overlap rate. The differences in RMSE between methods are small, never greater than 0.015. The RMSE for the item-eligibility method and for the MRM method with an acceleration parameter equal to zero are the same. Accuracy correlates negatively with the acceleration parameter.

The differences in bias can be seen in Figure 12. As in Study 1, the differences between methods are mainly found in negative trait levels. The higher the acceleration parameter, the greater is the bias. The bias of the item-eligibility method and the bias of the MRM method with an acceleration parameter equal to 0 are indistinguishable. The same results are found for RMSE, which is shown in Figure 13.

The results found with eCAT (Olea et al., 2004) are mainly the same as those obtained with randomly generated item banks in Study 1. Compared with defining just a single maximum exposure rate, defining multiple exposure rates considerably improves the security of the item bank with either no or only minor decrements in measurement accuracy.

Conclusions

As noted above, if the examinees know some of the items before they take the test, the validity of the test is adversely affected. To reduce the occurrence of this problem, it is important to reduce the variance of item exposure rates and thus the overlap rate between examinees (Chen et al., 2003).

The approach most widely used in CAT has been to impose a maximum exposure rate that no item should exceed. To do so, control parameters are calculated that determine the probability of an item's being administered once it has been selected, or the probability of its being eligible. Various methods have been proposed for calculating control parameters (Revuelta & Ponsoda, 1998; Sympson & Hetter, 1985; van der Linden & Veldkamp, 2004). However, several problems arise with this form of improved test security. First, although these methods are effective in eliminating overexposure, they have almost no impact on increased usage of underexposed items. Second, with these methods, when the maximum Fisher information selection rule is used, the quality of items selected decreases as the test progresses (Revuelta & Ponsoda, 1998).

This article presents an option for controlling exposure rates. Rather than defining only a single exposure rate as a limit, as many maximum exposure rates as items will be administered are marked. At the beginning of the test, the maximum exposure rate will be close to the minimum possible, increasing as the test progresses. The approach proposed involves only small modifications to the other method described, the item-eligibility method. A comparison of the performance of this new method with the item-eligibility method reveals that MRM clearly improves item bank security. Moreover, for the values of the acceleration parameter tested there is no relevant difference in accuracy between the MRM and the item-eligibility method. With randomly generated item banks, the overlap rate obtained with the item-eligibility method could be reduced by 40% while increasing RMSE by less than 0.01, when the acceleration parameter was set at 2. With eCAT, if an increment of 0.01 in RMSE is considered tolerable, the reduction in the overlap rate is 27%, with an acceleration parameter equal to 2.

Given the advantages of the MRM method, it can be considered the advisable option for controlling maximum exposure rates in CATs as it involves a more balanced usage of the item bank and delays possible distortion of trait estimation due to security problems with either no or only slight decrements in measurement accuracy.

References

Barrada, J. R., Olea, J., & Ponsoda, V. (2007). Methods for restricting maximum exposure rate in computerized adaptative testing. *Methodology*, *3*, 14-23.

Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2006). Estrategias de selección de ítems en un test adaptativo informatizado para la evaluación del inglés escrito [Item selection rules in a computerized adaptive test for the assessment of written English]. *Psicothema*, *18*, 828-834.

Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (in press). Incorporating randomness to the Fisher information for improving item exposure control in CATs. *British Journal of Mathematical and Statistical Psychology.*

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*, 431-444.

Chang, H. H. (2004). Understanding computerized adaptive testing—From Robbins-Monro to Lord and beyond. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 117-133). Thousand Oaks, CA: Sage.

Chang, H. H., & Ying, Z. (1999). α-Stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, *23*, 211-222.

Chang, H. H., & Ying, Z. (in press). To weight or not to weight? Balancing influence of initial items in adaptive testing. *Psychometrika.*

Chen, S. Y., Ankenmann, R. D., & Spray, J. A. (2003). The relationship between item exposure and test overlap in computerized adaptive testing. *Journal of Educational Measurement*, *40*, 129-145.

Chen, S. Y., & Doong, S. H. (2008). Predicting item exposure parameters in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, *61*, 75-91.

Chen, S. Y., Lei, P. W., & Liao, W. H. (in press). Controlling item exposure and test overlap on the fly in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*.

Hau, K. T., & Chang, H. H. (2001). Item selection in computerized adaptive testing: Should more discriminating items be used first? *Journal of Educational Measurement*, *38*, 249-266.

Li, Y. H., & Schafer, W. D. (2005). Increasing the homogeneity of CAT's item-exposure rates by minimizing or maximizing varied target functions while assembling shadow tests. *Journal of Educational Measurement*, *42*, 245-269.

Olea, J., Abad, F. J., Ponsoda, V., & Ximénez, M. C. (2004). Un test adaptativo informatizado para evaluar el conocimiento de inglés escrito: Diseño y comprobaciones psicométricas [A computerized adaptive test for the assessment of written English: Design and psychometric properties]. *Psicothema*, *16*, 519-525.

Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, *35*, 311-327.

Segall, D. O. (2004). Computerized adaptive testing. In K. Kempf-Lenard (Ed.), *The Encyclopedia of Social Measurement* (pp. 429-438). San Diego, CA: Academic Press.

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

Sympson, J. B., & Hetter, R. D. (1985, October). Controlling item exposure rates in computerized adaptive testing. In *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. (2003). Some alternatives to Sympson-Hetter item-exposure control in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, *28*, 249-265.

van der Linden, W. J. (2005). A comparison of item-selection methods for adaptive tests with content constraints. *Journal of Educational Measurement, 42,* 283-302.

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.

van der Linden, W. J., & Veldkamp, B. P. (2007). Conditional item-exposure control in adaptive testing using item-ineligibility probabilities. *Journal of Educational and Behavioral Statistics*, *32*, 398-418.

Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice*, *17*, 17-27.

Acknowledgments

This research was partly supported by a grant from the Spanish Ministry of Education and Science (project SEJ2006-08313/PSIC).

Author's Address

Address correspondence to Juan Ramón Barrada, Facultad de Psicología, Universidad Autónoma de Barcelona, 08193 Bellaterra, Spain; e-mail: juanramon .barrada@uab.es.