Psicothema 2009. Vol. 21, nº 2, pp. 313-320
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG

METODOLOGÍA

Comparison of methods for controlling maximum exposure rates in computerized adaptive testing

Juan Ramón Barrada, Francisco José Abad* and Bernard P. Veldkamp**
Universidad Autónoma de Barcelona, * Universidad Autónoma de Madrid and ** University of Twente

This paper has two objectives: (a) to provide a clear description of three methods for controlling the maximum exposure rate in computerized adaptive testing —the Symson-Hetter method, the restricted method, and the item-eligibility method— showing how all three can be interpreted as methods for constructing the variable sub-bank of items from which each examinee receives the items in his or her test; (b) to indicate the theoretical and empirical limitations of each method and to compare their performance. With the three methods, we obtained basically indistinguishable results in overlap rate and RMSE (differences in the third decimal place). The restricted method is the best method for controlling exposure rate, followed by the item-eligibility method. The worst method is the Sympson-Hetter method. The restricted method presents problems of sequential overlap rate. Our advice is to use the item-eligibility method, as it saves time and satisfies the goals of restricting maximum exposure.

Comparación de métodos para el control de tasa máxima en tests adaptativos informatizados. Este artículo tiene dos objetivos: (a) ofrecer una descripción clara de tres métodos para el control de la tasa máxima en tests adaptativos informatizados, el método Symson-Hetter, el método restringido y el método de elegibilidad del ítem, mostrando cómo todos ellos pueden interpretarse como métodos para la construcción del subbanco de ítems variable, del cual cada examinado recibe los ítems de su test; (b) señalar las limitaciones teóricas y empíricas de cada método y comparar sus resultados. Se obtienen resultados básicamente indistinguibles en tasa de solapamiento y RMSE con los tres métodos (diferencias en la tercera posición decimal). El método restringido es el mejor en el control de la tasa de exposición, seguido por el método de elegibilidad del ítem. El peor es el método Sympson-Hetter. El método restringido presenta un problema de solapamiento secuencial. Nuestra recomendación sería utilizar el método de elegibilidad del ítem, puesto que ahorra tiempo y satisface los objetivos de limitar la tasa máxima de exposición.

Experience and psychometrics tell us that not all items fit all examinees and not all items have the same quality. In the most common implementation of computerized adaptive testing (CAT; van der Linden & Pashley, 2000), this is taken into account, as the items administered to the examinees are those maximally informative for the estimated trait level until that moment, given the responses to the previous items. When this procedure is applied, some items are presented to a large proportion of the examinees. In this context, an examinee interested in ‘inflating’ his score could try to contact people previously examined, checking if they share with him the content of the items they received. The seriousness of this risk to the validity of the test will be greater the greater the test overlap rate is. The overlap rate is the mean proportion of items shared between examinees (Way, 1998).

Several methods have been proposed to reduce the overlap rate, all of which try to improve item bank security while retaining the accuracy of the test. The most common approach is to impose a maximum exposure rate that no item should surpass (Revuelta & Ponsoda, 1998; Sympson & Hetter, 1985; van der Linden & Veldkamp, 2004). In this article, we will focus on three different methods for controlling maximum exposure rates in CATs, showing their rationale, the points where they converge and their differences and, by means of a simulation study, comparing their performance.

Fecha recepción: 23-5-08 • Fecha aceptación: 9-12-08
Correspondencia: Juan Ramón Barrada
Facultad de Psicología
Universidad Autónoma de Barcelona
08193 Barcelona (Spain)
E-mail: juanramon.barrada@uab.es

Methods of controlling maximum exposure rate

Sympson-Hetter method

The Sympson and Hetter proposal (1985; Hetter & Sympson, 1997) has become the most commonly used method for item exposure control in CATs (van der Linden, 2003). This method is based on two different events for each item of the bank: (1) the item i is selected by the item selection rule (S_i); (2) the item i is administered (A_i). As an item cannot be administered if it has not been selected, it holds that

$$A_i \subset S_i \tag{1}$$

for each i. Thus,

$$P(A_i) \le P(S_i) \tag{2}$$

and

$$P(A_i) = P(A_i|S_i)P(S_i) \tag{3}$$

The goal of the Sympson-Hetter method (SH) is to set all the item exposure rates below a maximum exposure rate, r^{max}, fixed beforehand by the testing agency:

$$P(A_i) \le r^{max}, \tag{4}$$

holding that $Q/n \le r^{\max} \le 1$ (Chen, Ankenmann, & Spray, 2003), Q being the number of items to be administered and n the item bank size.

The $P(S_i)$ probabilities depend on the item selection rule applied, on the item bank composition and on the trait level distribution of the examinee population. As all these elements are fixed by design, to satisfy Equation 4, the only elements that can be manipulated in Equation 3 are the $P(A_i | S_i)$ probabilities.

Once an item is selected for an examinee, a random number belonging to the uniform interval (0, 1) is generated, and only if that number is smaller than $P(A_i | S_i)$ is that item administered. Otherwise, the item is not administered and is marked as non-selectable from then on for that examinee.

Suitable values $P(A_i | S_i)$ for reaching the goal are calculated through a series of iterative adjustments. Let us call t a step of this process. If in step t the probability $P^{(t)}(S_i)$ was lower than r^{max}, the probability $P^{(t+1)}(A_i)$ for step $(t+1)$ can be $P^{(t)}(S_i)$, as the exposure rate of the item is already lower than r^{max}. In the case that $P^{(t)}(S_i)$ is greater than r^{max}, we want $P^{(t+1)}(A_i)$ to be equal to r^{max}. This can be seen in Equation 5.

$$P^{(t+1)}(A_i) = \begin{cases} P^{(t)}(S_i) & \text{if} \quad P^{(t)}(S_i) \le r^{\max} \\ r^{\max} & \text{if} \quad P^{(t)}(S_i) > r^{\max} \end{cases} \tag{5}$$

Following Equation 3, and with some easy substitutions we can obtain the values for $P(A_i | S_i)$:

$$P^{(t+1)}(A_i|S_i) = \begin{cases} P^{(t)}(S_i)/P^{(t+1)}(S_i) & \text{if} \quad P^{(t)}(S_i) \le r^{\max} \\ r^{\max}/P^{(t+1)}(S_i) & \text{if} \quad P^{(t)}(S_i) > r^{\max} \end{cases} \tag{6}$$

Before the simulation of the $(t+1)$-th cycle, as we do not have the value of $P^{(t+1)}(S_i)$, we employ as an estimation of $P^{(t+1)}(S_i)$ the value of $P^{(t)}(S_i)$:

$$\hat{P}^{(t+1)}(S_i) = P^{(t)}(S_i) \tag{7}$$

Rewriting Equation 6, we obtain:

$$P^{(t+1)}(A_i|S_i) = \begin{cases} 1 & \text{if} \quad P^{(t)}(S_i) \le r^{\max} \\ r^{\max}/P^{(t)}(S_i) & \text{if} \quad P^{(t)}(S_i) > r^{\max} \end{cases} \tag{8}$$

Equation 8 is the usual way of formulating the SH method. Another possible way involves replacing $P^{(t)}(S_i)$ by $P^{(t)}(A_i)/P^{(t)}(A_i | S_i)$.

$$P^{(t+1)}(A_i|S_i) = \begin{cases} 1 & \text{if} \quad P^{(t)}(A_i)/P^{(t)}(A_i|S_i) \le r^{\max} \\ r^{\max} P^{(t)}(A_i|S_i)/P^{(t)}(A_i) & \text{if} \quad P^{(t)}(A_i)/P^{(t)}(A_i|S_i) > r^{\max} \end{cases} \tag{9}$$

As many iterations as needed are calculated until the maximum exposure rate is stabilized.

A different formulation of the SH method

Another way of defining the SH method will allow us to show the relationship between this method and the other alternatives that have been proposed for limiting item overexposure. In this formulation, two events are defined: (1) item i is eligible for the examinee (E_i); (2) item i is administered (A_i). In this case, the exposure control is achieved by restricting the proportion of examinees for which an item can be eligible. For each candidate, a subset of eligible items is formed before any item has been administered. During the administration, only items from this subset can be administered. As any administered item has to be eligible, it holds that

$$A_i \subset E_i \tag{10}$$

and

$$P(A_i) \le P(A_i|E_i) \tag{11}$$

Thus,

$$P(A_i) = P(A_i|E_i)P(E_i) \tag{12}$$

Again, $P(A_i)$ are the values that we wish to control in Equation 4. In this case, $P(A_i | E_i)$ is fixed by design and $P(E_i)$ is the control parameter that we manipulate for achieving our goal.

According to this approach, before testing an examinee, a random number from (0, 1) is generated for each item in the bank. Only if the number is smaller than $P(E_i)$, can that item be administered.

Following a logic similar to that applied above, and using the values of $P^{(t)}(A_i | E_i)$ as estimations of $\hat{P}^{(t+1)}(A_i | E_i)$, we can show the equation needed for calculating the $P(E_i)$ values:

$$P^{(t+1)}(E_i) = \begin{cases} 1 & \text{if} \quad P^{(t)}(A_i)/P^{(t)}(E_i) \le r^{\max} \\ r^{\max} P^{(t)}(E_i)/P^{(t)}(A_i) & \text{if} \quad P^{(t)}(A_i)/P^{(t)}(E_i) > r^{\max} \end{cases} \tag{13}$$

Comparing Equation 9 with Equation 13, it can be seen that the $P(A_i | S_i)$ control parameters and the $P(E_i)$ control parameters lead to the same set of values. The two methods differ only theoretically. With the $P(A_i | S_i)$ parameters, exposure control is done after item selection. With the $P(E_i)$ parameters, exposure control is done before item selection.

The time needed for calculating the control parameters and the time required for item selection when the CAT is operative should be taken into account when comparing both approaches. When the $P(A_i | S_i)$ parameters are used, the time needed for item selection is, basically, the time for evaluating all the items of the bank according to the item selection rule. When $P(E_i)$ is used as the control parameter, we have to construct the eligible sub-bank (a trivial operation for modern computers) and to evaluate just this subset according to the item selection rule. As the computation time required for evaluting the item selection rule increases, the time saved with the second formulation of the SH method will be greater. For instance, the the item selection rules based in the Kullback-Leibler function (Chang & Ying, 1996) are computationally more demanding that the selection by means of maximum Fisher information in the estimated trait level. Also, reducing the value of r^{max} implies reducing the size of the set of eligible items and, thus, the number of items to be evaluated when the $P(E_i)$ parameters are in use. So, the more severe the restrictions on maximum exposure rate, the greater the difference in time required by both approaches.

Another advantage of the second formulation of the SH method is that, in this way, the SH method can be easily combined with the shadow test method (van der Linden & Reese, 1998), one of the best alternatives for incorporating content constraints in CATs (van der Linden, 2005).

To compare these two formulations of the SH method with other exposure control methods, the control parameters will be called k_i parameters from now on.

Limitations of the SH method

As Barrada, Olea and Ponsoda (2007) and van der Linden (2003) have noted, the SH method presents several limitations. Firstly, this method is unable to set all the exposure rates equal to or below the desired value. The maximum exposure rate is slightly over r^{max}. As the values of $P(S_i)$ or of $P(A_i | E_i)$ are not kept constant from cycle to cycle, the estimations used for calculating the k_i parameters differ from the empirical values, so that the restriction in Equation 4 cannot be satisfied.

Second, the k_i parameters calculated are dependent on the distribution of the estimated trait levels and on the item bank. Any change in either means that the process of calculation of the control parameters needs to be repeated (Chang & Harris, 2002). The change in the distribution of the examinees' estimated trait level may be due to a change in the distribution of the real trait levels (differences in the academic curriculum, for example, if academic abilities are being assessed), or to alterations of the estimations distribution unrelated to changes in the distribution of real trait levels (some examinees knowing a part of the item bank in advance would increase the mean of the estimated trait levels). The composition of the bank changes whenever an item is removed from the item bank or a new item is incorporated, tasks which are necessary for the maintenance of the item bank.

Another limitation is the simulation process necessary to obtain the k_i parameters and the time consumed by this process. The time needed for calculating the final k_i values is dependent on the number of cycles that has to be simulated and the time to compute each cycle. The number of cycles depends on, among other elements, the degree of exposure control desired (Barrada et al., 2007). The lower the r^{max} value, the more cycles are needed, as more k_i parameters differ from 1. According to van der Linden (2003), 10-12 cycles may be sufficient to calculate the final k_i parameters. If the SH method is applied conditional upon trait levels, real (Stocking & Lewis, 1998) or estimated (Stocking & Lewis, 2000), this figure must be multiplied by the number of levels used.

The time needed for each cycle depends on the number of examinees simulated. Stocking and Lewis (1998) have shown that the greater this number, the closer the maximum exposure rate will be to r^{max}. The time needed for each cycle also depends on the item selection rule used. The one that seems most accurate, the Kullback-Leibler information function weighted by the likelihood function (Barrada, Olea, Ponsoda, & Abad, 2009; Chang & Ying, 1996; Chen, Ankenmann, & Chang, 2000), is also one of the slowest to compute and, thus, to converge.

In recent years, several modifications of the SH method have been proposed to accelerate estimation of the k_i parameters (Barrada et al., 2007; Chen & Doong, 2008; van der Linden, 2003). In these, it is still necessary to use an iterative simulation process. Other approaches do not require any prior simulation, but rather adapt the k_i parameters for each examinee as the test goes on (Revuelta & Ponsoda, 1998; van der Linden & Veldkamp, 2004). We shall now describe these methods.

Restricted method

The restricted method (RT; Revuelta & Ponsoda, 1998) adapts the subset of the bank which is available for administration for each examinee. The control parameters can adopt just two values, 0 and 1. The k_i parameter will be set at 0 if the exposure rate of the item until the j-th examinee is greater than or equal to r^{max}; otherwise, the control parameter will be set at 1:

$$k_i^{(j+1)} = \begin{cases} 1 & if \quad P^{(1\ldots j)}(A_i) < r^{\max} \\ 0 & if \quad P^{(1\ldots j)}(A_i) \geq r^{\max} \end{cases} \tag{14}$$

In the original proposal (Revuelta & Ponsoda, 1998), the k_i parameters were used to define the sub-bank of items available for administration (so, as $P(E_i)$ parameters). Barrada et al. (2007) used the k_i parameters of the restricted method to determine if the item could be administered after it was selected (so, as $P(A_i | S_i)$ parameters).

The RT method has at least three advantages over the SH method: (a) the k_i parameters are adapted on-the-fly, saving computation time; (b) the restriction imposed by Equation 4 is met to a greater extent; and (c) the exposure rates of the items do not surpass r^{max}, even when there are changes in the item bank composition or in the trait level distribution. There is a price to pay for these advantages however: a slight reduction in measurement accuracy when compared with the SH method (Revuelta & Ponsoda, 1998).

A problem with the RT method was been described by Chen, Lei and Liao (2008): the predictability of administration sequence of some items. Consider the case of an item with an exposure rate, if no restriction is applied, equal to 1. The sequence of k_i values for this item when the value of r^{max} is set at 0.25 can be seen in Table 1. This item will be administered to the first examinee and will not be eligible again until the 6th examinee. After that point, that item will always follow the same pattern: eligible once, not eligible three times. In this situation, the overlap between every four examinees should clearly be greater than the overlap in the overall sample of examinees.

Recently, a new method that combines parts of the SH method (probabilistic method) and of the RT method (constant update of the k_i parameters) has been proposed.

Item-eligibility method

The item-eligibility method (IE; van der Linden & Veldkamp, 2004, 2007), formulated in Equation 15, clearly resembles the SH method (Equations 9 and 13). As in the restricted method, the k_i parameters are adjusted for each new examinee. The parameters for the *(j+1)*-th examinee are calculated using the exposure rates from when the test starts to the *j*-th examinee:

$$k_i^{(j+1)} = \begin{cases} 1 & \text{if} \quad P^{(1\ldots j)}(A_i)/k_i^{(j)} \le r^{\max} \\ r^{\max} k_i^{(j)} / P^{(1\ldots j)}(A_i) & \text{if} \quad P^{(1\ldots j)}(A_i)/k_i^{(j)} > r^{\max} \end{cases} \tag{15}$$

In this method, the k_i parameters are $P(E_i)$ parameters. While in the SH method the values for the k_i parameters belonged to the interval [r^{max}, 1] and in the RT method the possible values were just {0, 1}, in the IE method the interval in which the k_i parameters need to be placed is (0, 1]. Prior to the administration of any item to an examinee, a random number belonging to the uniform interval (0, 1) is generated for each item and, only if that number is smaller than the k_i parameter, does that item belong to the sub-bank of eligible items.

The main advantage of the IE method over the SH method is that no time-consuming simulation studies are necessary to find admissible values for control parameters k_i of the items. Instead, the method can be implemented on-the-fly during operational testing. The values for the control parameters k_i are automatically adapted, based on the control parameters and probability of administration of the items.

Unlike the RT method, the IE method is probabilistic in nature. Because of this, the maximum exposure rate might be slightly violated for some of the most popular items. On the other hand, the eligible subset that is selected for a candidate only depends by chance on the previously administered items. Because of this, problems due to the predictability of the administration sequence are not expected to occur. In Table 2, an application of the IE method for two items and 10 test administrations is presented.

This example tries to reflect to probabilistic nature of the IE method and how the control parameters are updated. Item 1 is administered to the first examinee and is not administered again until the tenth. Item 2 is administered to the first two examinees. It can be seen how the same probabilities of administration do not lead to the same values for the control parameter and how the values of the k_i parameter can reach values very near 0.

We have described three methods for controlling maximum exposure rate in CATs. All set different exposure control parameters to define the probability that each item of the bank belong to the sub-set of items that can be administered to each examinee. Revuelta and Ponsoda (1998) compared the SH and the RT method, finding that the RT performed better in controlling the exposure rates, but at the cost of a very small increase in measurement error. Van der Linden and Veldkamp (2004) considered that the SH method could not be used in combination with the shadow test approach, so they did not compare the IE method with the SH method. However, they did not consider the implementation of the SH method proposed in Equation 13. As far as we know, no study has been presented comparing the three methods. We now present a simulation study with these data.

SIMULATION STUDY

Method

Ten item banks were generated, each with 500 items, with parameters *a*, *b*, and *c* taken at random from distributions N(1.2, 0.25), N(0, 1), and N(0.25, 0.02), respectively. Length of the CAT was set at 25 items. Examinees' trait level was taken at random from a population N(0, 1). Initial trait level was extracted at random within the interval (–0.5, 0.5). The number of examinees simulated was 5000 per condition.

Two different methods of trait level estimation were used. The first method was maximum-likelihood (Birnbaum, 1968). Maximum-likelihood estimation has no solution in real numbers

Table 1
Example of an application of the RT method with r^{max} equal to 0.25

j	administered	$p^{(1\ldots j)}(A)$	$k^{(j)}$
1	Yes	1	1
2	No	0.5000	0
3	No	0.3333	0
4	No	0.2500	0
5	No	0.2000	0
6	Yes	0.3333	1
7	No	0.2857	0
8	No	0.2500	0
9	No	0.2222	0
10	Yes	0.3000	1

Table 2
Example of an application of the IE method with r^{max} equal to 0.25 for two different items

	item 1			item 2		
j	administered	$p^{(1\ldots j)}(A)$	$k^{(j)}$	administered	$p^{(1\ldots j)}(A)$	$k^{(j)}$
1	Yes	1	1	Yes	1	1
2	No	0.5000	0.2500	Yes	1	0.25
3	No	0.3333	0.1250	No	0.6667	0.0625
4	No	0.2500	0.0938	No	0.5000	0.0234
5	No	0.2000	0.0938	No	0.4000	0.0117
6	No	0.1667	0.1172	No	0.3333	0.0073
7	No	0.1429	0.1758	No	0.2857	0.0055
8	No	0.1250	0.3076	No	0.2500	0.0048
9	No	0.1111	0.6152	No	0.2222	0.0048
10	Yes	0.2000	1	No	0.2000	0.0054

when there is a constant response pattern, all correct or all incorrect responses. Therefore, until there was at least one correct and one incorrect response or the test was finished, θ was assigned using the method proposed by Dodd (1990). When all the responses were correct, $\hat{\theta}$ was increased by $(b_{max} - \hat{\theta})/2$. If all the responses were incorrect, $\hat{\theta}$ was reduced by $(\hat{\theta} - b_{min})/2$. In these formulae b_{max} and b_{min} refer to the highest and lowest b parameters, respectively, of the entire item bank. If the trait level estimation was greater than 4, it was fixed at 4. If it was lower than -4, it was fixed at -4. The second method was expected a posteriori (EAP; Bock & Mislevy, 1982), with a uniform prior over [-4,4]. The item selected was the one that maximized the Fisher information function at the estimated trait level.

Typically, r^{max} is chosen to be in the range of 0.2 to 0.3 (van der Linden & Veldkamp, 2007). Two different values of r^{max} were used: 0.25, in the range of common values, and 0.15, slightly more stringent than the values above. We also simulated the condition without restriction in r^{max} (r^{max}= 1).

The k_i parameters used in the SH method were those obtained in the 25th cycle. For the SH, RT and IE methods, the k_i parameters were considered as $P(E_i)$ parameters: the control parameters defined the probability that each item belonged to the sub-bank of eligible items.

Five variables were used for the comparison between methods: (a) observed maximum exposure rates; (b) proportion of items with exposure rates over r^{max}; (c) mean exposure of items with exposure over r^{max}; (d) overlap rate, calculated following Equation 16; and (e) RMSE, calculated with Equation 17.

The overlap rate was:

$$\hat{T} = \frac{n}{Q} S^2_{P(A)} + \frac{Q}{n} \tag{16}$$

where $\hat{T}$ is the large-sample approximation of the overlap rate (Chen et al., 2003), Q is the test length, n the item bank size and $S^2_{P(A)}$ is the variance of the item exposure rates.

The RMSE was:

$$RMSE = \left(\sum_{g=1}^{m} \left(\hat{\theta}_g - \theta_g \right)^2 / m \right)^{1/2} \tag{17}$$

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

Results

We will describe the results according to the different variables manipulated, r^{max}, estimation method and method for controlling maximum exposure rate. The results of these variables can be seen in Table 3. We also include a point describing the sequential test overlap, one of the possible problems of the RT method (Chen et al., 2008), but only for the condition of r^{max} equal to 0.25 (the condition in which this sequence could be most clearly detected). Finally, we describe, for an item of our item banks, how the k_i parameter is updated with the IE method.

Effects of r^{max}: as could be expected, reducing r^{max} from 0.25 to 0.15 reduces the maximum observed exposure rate and the mean exposure of items with exposure over r^{max}. Lowering the value of r^{max} implies that more items will have their exposure controlled by the k_i parameters, more k_i parameters will be different from 1, and so, in the SH and IE methods, the more frequently the eligibility of an item is determined by a random number. Reducing r^{max} increases the proportion of items with exposure rate over this limit because the randomness of the selection increases as the r^{max} decreases. The lower the value of r^{max}, the lower the overlap rate. With r^{max} equal to 0.25, to examinees share, on average, 4.86 items (out of 25). With r^{max} set at 0.15, they share 3.31. This reduction of the overlap by 68.2% is achieved at the cost of an increase in RMSE of .01. This pattern of results are coherent with previous results (Barrada, Olea, Abad, 2007; Barrada et al., 2007). The condition without restrictions on the exposure rates was the one

Table 3
Maximum exposure rate, proportion of items with exposure rate over r^{max}, mean exposure of items with exposure rate over r^{max}, overlap rate and RMSE according to r^{max}, trait level estimation method and method for controlling maximum exposure rate

r^{max}	estimation	method	maximum exposure rate	proportion of items with exposure over r^{max}	mean exposure of items with exposure over r^{max}	overlap rate	RMSE
1	ML	–	0.7483	–	–	0.2730	0.2535
	EAP	–	0.7789	–	–	0.2926	0.2437
0.25	ML	SH	0.2676	0.0466	0.2559	0.1942	0.2644
		RT	0.2500	0	–	0.1929	0.2647
		IE	0.2538	0.0324	0.2513	0.1933	0.2649
	EAP	SH	0.2676	0.0478	0.2563	0.1957	0.2510
		RT	0.2500	0	–	0.1944	0.2521
		IE	0.2542	0.0366	0.2515	0.1952	0.2533
0.15	ML	SH	0.1657	0.1144	0.1553	0.1338	0.2713
		RT	0.1500	0	–	0.1300	0.2771
		IE	0.1539	0.0998	0.1513	0.1325	0.2736
	EAP	SH	0.1682	0.1290	0.1551	0.1347	0.2616
		RT	0.1500	0	—	0.1309	0.2630
		IE	0.1544	0.0990	0.1514	0.1331	0.2608

with greatest maximum exposure rate, although not reaching 1, greater overlap rate and lower RMSE.

Effects of the trait level estimation method: when r^{max} was lower than 1, both maximum-likelihood estimation and EAP estimation had the same results (differences found in the third decimal place) in all the variables considered, with the exception of the RMSE. The RMSE with the EAP estimation was consistently lower than the RSME with maximum-likelihood estimation, a difference of .01. This implies that it is possible to obtain basically the same RMSE with EAP estimation and r^{max} equal to 0.15 as with maximum-likelihood estimation and r^{max} set at 0.25. When r^{max} was equal to 1, EAP estimation produced a higher maximum exposure rate and overlap rate, and, again, lower RMSE.

Effects of the methods for controlling the maximum exposure rate: the performance of the different methods was not modulated by the other variables considered in this study, so we will speak about the overall means. The lowest maximum exposure rate was obtained with the RT method. Actually, with this method the maximum exposure rate was equal to r^{max}, so no items had an exposure rate over r^{max}. The IE method was the second best method in satisfying the restriction of r^{max}. The maximum exposure rate with the SH method was .017 over the limit, so it was the method with the worst results in this indicator. The IE method outperformed the SH method in the proportion of items with exposure over r^{max}. The difference in the mean exposure of these items, when comparing the SH and the IE methods, can be considered as negligible. Following these results, the lowest overlap rate was found with the RT method, next with the IE method and the method with greatest overlap was the SH method, although this differences can be considered as practically irrelevant. The order in the results of the RMSE is the reverse of that in the overlap rate, but, again, the differences between methods are minimal.

Sequential test overlap: the main results on this point, for the condition with r^{max} equal to 0.25, are shown in Table 4. There, the overlap and maximum exposure rate considering every four examinees can be seen. One condition is for examinees 1-5-9-…, the next for examinees 2-6-10-…, and so on. It should be noted that, when no restrictions on r^{max} were imposed, no item had an exposure rate equal to 1. In these conditions, both for the SH and the IE, the overlap rates are, basically, the same than when considering the whole sample of examinees. The maximum exposure rates, when considering the items administered to every four examinees, are slightly over the maximum exposure rates when calculated with the complete set of examinees, probably because of lower sample size in the former condition. These results clearly change with the RT method. While in the whole sample the RT method fixed the maximum exposure rate equal to r^{max}, when considering every four examinees the maximum exposure rate is clearly over this limit. In addition, the overlap rate is greater when considering every four examinees than when considering all the examinees simulated. In concordance with what we described previously and shown in Table 1, the greatest overlap and greatest maximum exposure rate is with examinees 2-6-10-… For examinees 2-6-10…, the maximum exposure rate and overlap rate are the same than those obtained when no restriction on r^{max} was imposed (compare with Table 3). In Table 1 we have shown that, in the case of r^{max} equal to 0.25, examinees in position $(2+4h)$-th in the sequence of examinees (h belonging to the natural numbers) will have available for administration the items with an exposure rate equal to 1 if no restriction on r^{max} was applied. This could be generalized to the rest of the items of the bank. What is the same, for examinees in those positions, the sub-bank of eligible items is the same than the whole item bank.

Updating of the k_i parameter in the IE method: we have shown that the IE method can easily accommodate the restriction imposed in r^{max}. Given the similarities between Equations 13 and 15, it could be considered that the k_i parameters calculated with the IE method, after a large number of examinees, stabilize and converge with the k_i parameters obtained with the SH method. In other words, after a large number of examinees, there is no reason to continue updating the k_i parameters for the IE method. With an example, we will show that this is not a correct interpretation of how the IE method works.

We have selected the item with maximum exposure rate in the first item bank simulated in the condition of r^{max} equal to 0.25 and ML estimation. We have plotted the probability of administration and the value of the control parameter after each examinee, for both of the values of r^{max} simulated. This graph is presented in Figure 1. There, it can be seen how the exposure rate for this item stabilizes around r^{max} after no more than 1000 examinees, although some minor oscillations can be detected. In addition, when the exposure rate is basically stable, the k_i parameter oscillates throughout the 5000 examinees. For some subsets of examinees the probability of eligibility for that item is quite small (markedly below r^{max}), while for other subsets this probability is much less restrictive. We have verified that this result is not specific to the item selected for the plot. Given these results, it is clear that the updating of the k_i parameters should never be stopped with the IE method.

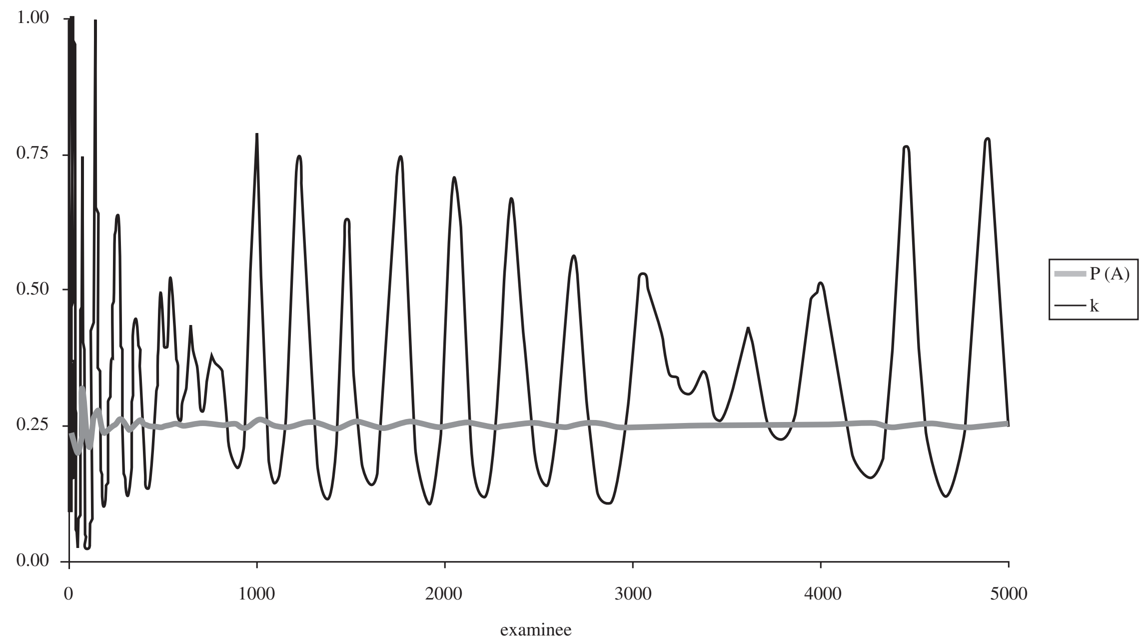
Table 4
Maximum exposure rate and overlap rate according to trait level estimation method and method for controlling maximum exposure rate for every four examinees

estimation	method	maximum exposure rate				overlap rate			
		1-5-9-…	2-6-10-…	3-7-11-…	4-8-12-…	1-5-9-…	2-6-10-…	3-7-11-…	4-8-12-…
ML	SH	0.2760	0.2771	0.2794	0.2769	0.1943	0.1948	0.1942	0.1952
	RT	0.3261	0.7474	0.3327	0.2910	0.1872	0.2730	0.2043	0.1904
	IE	0.2753	0.2722	0.2734	0.2699	0.1944	0.1930	0.1950	0.1929
EAP	SH	0.2777	0.2828	0.2786	0.2805	0.1966	0.1962	0.1961	0.1957
	RT	0.3350	0.7726	0.3550	0.2932	0.1911	0.2909	0.2106	0.1918
	IE	0.2750	0.2730	0.2697	0.2714	0.1970	0.1959	0.1953	0.1946

Conclusions

This paper had four objectives: (a) to provide a clear description of three methods for controlling the maximum exposure rate in CATs, the SH method, the RT method and the IE method; (b) to indicate the limitations of each; (c) to show how all can be interpreted as methods for constructing the variable sub-bank of items from which each examinee receives the items in his test; (d) to compare their performance.

The first three goals have been addressed in the section where the methods were described. The final goal was addressed by means of a simulation study, where the three methods were compared in 5 variables. We found that different values of r^{max} or different methods for estimating the trait level did not change the pattern of results. As could be expected, RT, a deterministic method, is the only method that completely satisfies the restriction on maximum exposure rate imposed by r^{max}. The IE method, which adapts the k_i parameters on-



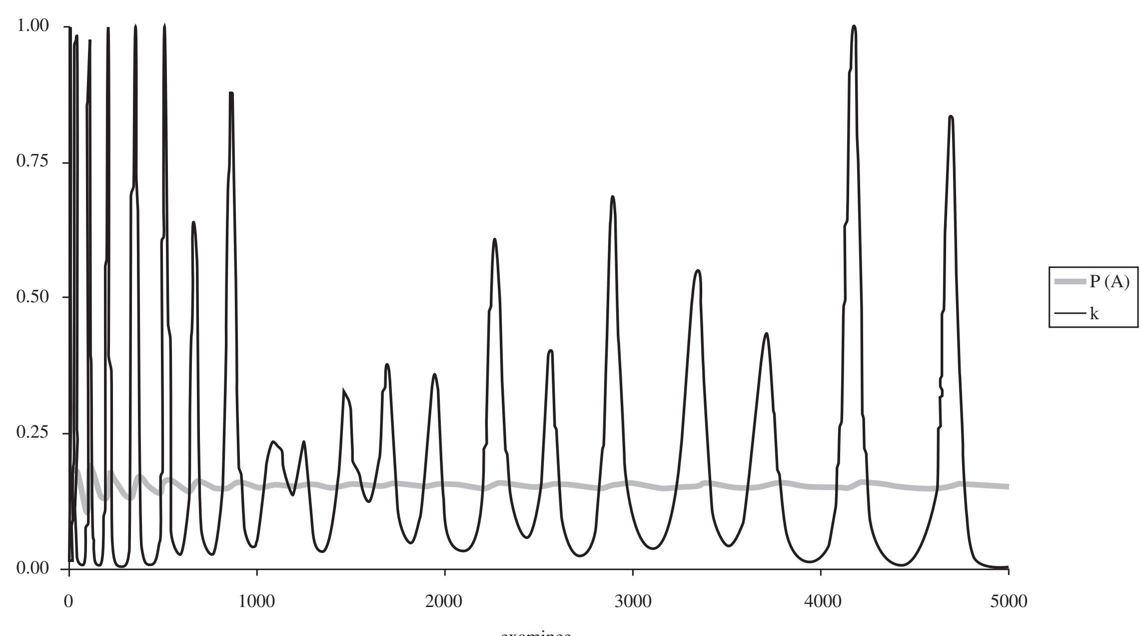


Figure 1. *Probability of administration and value of the control parameter for an item with the IE method according to examinee position. Top: r^{max} equal to 0.25. Bottom: r^{max} equal to 0.15*

the-fly, is the second best method in terms of satisfying the desired restrictions on the maximum exposure rate. The method with the worst performance was the SH method, because of the assumptions made (Equation 7; van der Linden, 2003). Despite these differences in the maximum exposure rate, the proportion of items with exposure over r^{max} and the mean exposure of items with exposure over r^{max}, all three methods offer negligible differences (in the third decimal place) in the overlap rate and the RMSE.

Chen et al. (2008) pointed out that a limitation of the RT method was the sequential overlap. We have replicated their results, showing that the overlap between examinees is not independent of the order in which they were tested. For the SH and the IE methods we do not find this problem.

Summarizing:

(a) The SH method was the one that most clearly surpassed the maximum exposure rate defined by r^{max}. It is important to note that the simulations have been carried out under optimal conditions for the SH method: the examinees' trait distribution used for defining the k_i parameters was exactly the same as the distribution used to calculate the final results. As we have said above, whenever there is a mismatch between both distributions the SH method will not be able to control the maximum exposure rate (Chen & Doong, 2008). The RT and IE methods, which adapt the k_i parameters on-the-fly, do not have this limitation.
(b) With the RT and IE methods we save the time needed to establish the k_i parameters.
(c) With the RT method we find a problem of sequential overlap that we do not find with the other methods.

Merging all this information, we consider that the preferred method is the IE method, as it is a method with none of the limitations described for the other methods and it is able to control the maximum exposure rate at a level almost equal to r^{max}. Despite this, the differences between methods are very small.

References

Barrada, J.R., Olea, J., & Abad, F.J. (2008). Rotating item banks versus restriction of maximum exposure rates in computerized adaptive testing. *Spanish Journal of Psychology*, *11*, 618-625.

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. (2009). Item selection rules in Computerized Adaptive Testing: Accuracy and security. *Methodology, 5,* 7-17.

Birnbaum, A. (1968). Some latent ability models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.): *Statistical theories of mental test scores* (pp. 392-479). Reading, MA: Addison-Wesley.

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., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement, 20*, 213-229.

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.

Chen, S.Y., Ankenmann, R.D., & Chang, H.H. (2000). A comparison of item selection rules at the early stages of computerized adaptive testing. *Applied Psychological Measurement, 24*, 241-255.

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. (2008). Controlling item exposure and test overlap on the fly in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology, 61*, 471-492.

Dodd, B.G. (1990). The effect of item selection procedure and stepsize on computerized adaptive attitude measurement using the rating scale model. *Applied Psychological Measurement, 14*, 355-366.

Hetter, R.D., & Sympson, J.B. (1997). Item exposure control 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.

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

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.

Stocking, M.L., & Lewis, C.L. (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). Dordrecht, The Netherlands: Kluwer Academic.

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 method for adaptive tests with content constraints. *Journal of Educational Measurement, 45*, 283-302.

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). Norwell, MA: Kluwer.

van der Linden, W.J., & Reese, L.M. (1998). A model for optimal constrained adaptive testing. *Applied Psychological Measurement, 22*, 259-270.

van der Linden, W.J., & Veldkamp, B.P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational & 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.