Methods for Restricting Maximum Exposure Rate in Computerized Adaptative Testing

Juan Ramon Barrada, Julio Olea, and Vicente Ponsoda

Universidad Autonoma de Madrid, Spain

Abstract. The Sympson-Hetter (1985) method provides a means of controlling maximum exposure rate of items in Computerized Adaptive Testing. Through a series of simulations, control parameters are set that mark the probability of administration of an item on being selected. This method presents two main problems: it requires a long computation time for calculating the parameters and the maximum exposure rate is slightly above the fixed limit. Van der Linden (2003) presented two alternatives which appear to solve both of the problems. The impact of these methods in the measurement accuracy has not been tested yet. We show how these methods over-restrict the exposure of some highly discriminating items and, thus, the accuracy is decreased. It also shown that, when the desired maximum exposure rate is near the minimum possible value, these methods offer an empirical maximum exposure rate clearly above the goal. A new method, based on the initial estimation of the probability of administration and the probability of selection of the items with the restricted method (Revuelta & Ponsoda, 1998), is presented in this paper. It can be used with the Sympson-Hetter method and with the two van der Linden's methods. This option, when used with Sympson-Hetter, speeds the convergence of the control parameters without decreasing the accuracy.

Keywords: computerized adaptive testing, item exposure control, Sympson-Hetter method, overlap rate, item bank security

One of the objectives of administering tests is accurate assessment of the examinee's trait level. In order to achieve adequate measurement, it is necessary that the probability of responding correctly to the items is marked solely by their psychometric characteristics and by the examinee's trait level. In the case of an examinee receiving the test with prior knowledge of the items to which he will have to respond, this would no longer hold and there would be an over-estimation of his trait level that would reduce the test's validity.

This risk is especially present when the test is applied by means of a computerized adaptive test (CAT). In this kind of test, the items in the item bank remain operative for a reasonably long period of time. This means that a future examinee can obtain knowledge of part of the item bank if he receives information from an examinee already tested who remembers the items he faced. The risk will be higher the higher the overlap rate between examinees, this being understood as the proportion of items shared, on average, by two examinees (Way, 1998).

Normally, the item-selection rule applied in CATs seeks the item not yet administered that maximizes measurement efficacy given the estimated trait level (van der Linden & Pashley, 2000). This means that, if no additional restriction is applied, a large proportion of the items in the bank are not presented to any examinee, while a few of them have a very high exposure rate. Such high variance in the exposure rate of the items in a bank implies a high overlap rate, as shown by Chen, Ankenmann, and Spray (2003).

In order to reduce this risk, various strategies have been proposed. Some seek mainly to increase the exposure rate of under-exposed items (Chang & Ying, 1999; Li & Schafer, 2005; Revuelta & Ponsoda, 1998). Others, however, tackle the problem of over-exposure, by forcing the maximum exposure rate of the items to be below a prespecified level, r_{max} (Revuelta & Ponsoda, 1998; Sympson & Hetter, 1985; van der Linden, 2003; van der Linden & Veldkamp, 2004). Combinations of the two approaches are also possible (Leung, Chang, & Hau, 2002; Revuelta & Ponsoda, 1998). Of these methods, that most commonly used is the Sympson-Hetter method (1985; Hetter & Sympson, 1997).

The Sympson-Hetter Method as a Means of Controlling Exposure

The Sympson-Hetter (SH) method is based in 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). This method provides a form of controlling the maximum exposure rate of items: it is sought to situate the maximum exposure rate of all the items equal or below r_{max}.

$$\max[P(A_i)] \leq r_{max} \tag{1}$$

As an item cannot by administered if it has not been selected, it holds that

DOI 10.1027/1614-1881.3.1.14

$$P(A_i) = P(A_i|S_i)P(S_i) \tag{2}$$

The $P(A_i|S_i)$ value, appropriately defined, would allow fulfillment of the criterion defined in Equation 1. Through a series of simulation cycles, it is attempted to establish the values of $P(A_i|S_i)$ so that they satisfy Equation 1 (from here on, we shall refer to the $P(A|S)$ as k parameters). The k parameters for the cycle $t + 1$ derive from making in Equation 2 $P(A_i)$ equal to r_{max} and setting the limitation that the maximum value of the k parameter can be 1.

$$k_i^{t+1} = \begin{cases} 1 & \text{if } P^{(t)}(S_i) \le r_{\max} \\ r_{\max}/P^{(t)}(S_i) & \text{if } P^{(t)}(S_i) > r_{\max} \end{cases} \tag{3}$$

When an item is selected, a random number belonging to U(0, 1) is generated, and only if this number is lower than $P(A_i|S_i)$ is that item administered. In the opposite case, the item is not administered and is marked as nonselectable for that examinee.

In normal practice, the SH method is applied for controlling over-exposure in the overall population and also conditioned the several trait levels, real (Stocking & Lewis, 1998) or estimated (Stocking & Lewis, 2000). This method has several limitations, which we shall now describe.

Limitations of the SH Method

The SH method raises two main problems. The first is related to the long time necessary for stabilizing the estimations of the k parameters. The second refers to the method's inability to guarantee that the condition established in Equation 1 is met.

1. Problems related to computation time: The computation time employed with the SH method increases, on the one hand, as the complexity of calculation of the different item selection rules increases, and, on the other, the lower the value set for r_{max}. In both cases, the seemingly advisable alternatives would involve a long computation time. That which seems the most accurate item-selection rule, the Kullback-Leibler function weighted by the likelihood function (Chang & Ying, 1996; Chen, Ankenmann, & Chang, 2000; Barrada, Olea, & Ponsoda, 2005), is also one slowest in its calculation. In the study by Barrada et al., this rule was around 50 times slower than that normally employed, which is selection of the item with maximum Fisher information for the estimated trait level. It is also important to note that Chang (2004) has shown the high risk involved in failing to impose strict control on item exposure in CATs, being, so, recommendable setting severe restrictions for r_{max}.
   The problem of computation time is exacerbated by two factors. First, if the SH method is applied conditioned to trait levels, the time employed is multiplied by the number of levels used, normally around 10 to 12 (van der Linden, 2003). Second, maintenance of an operative bank involves periodical removal of some of its items and the inclusion of new items. Each time the composition of the bank is changed, even by just one item, the k parameters have to be re-estimated (Chang & Harris, 2002).
2. Problems related to the convergence of exposure rates equal or below r_{max}: With the SH method, some items still have exposure rates over r_{max}. Van der Linden (2003) shows why it is impossible for this method to satisfy Equation 1. The $k_i^{(t+1)}$ parameters are calculated using $P^{(t)}(S_i)$ as the estimation of $\hat{P}^{(t+1)}(S_i)$. But the probability of selection of an item does not remain constant from cycle to cycle. If that were the case, the k parameters could be calculated with a single iteration and converge fully. The discrepancy between $P^{(t+1)}(S_i)$ and $\hat{P}^{(t+1)}(S_i)$ explains why the maximum exposure rate stabilizes slightly above r_{max}.

Alternatives to the SH Method

Van der Linden (2003) proposes various alternatives for a quicker and more effective calculation of the k parameters. Of these, two stand out as the most appropriate, judged on the basis of maximum exposure rate, number of items with a rate over r_{max} and mean rate of items over r_{max}. In all of these variables, the new methods are superior to the SH method.

The two methods share two characteristics that distinguish them from the SH method, and which, in combination, are what seem to make them more effective and efficient. First, the adjustments in the k parameter are only made when the item exposure rate is above r_{max}. Therefore, k parameters that manage to situate the exposure rate below the limit are not readjusted. In this way it is intended to avoid a situation whereby items that for one cycle had had an exposure rate below r_{max} went above it in the following cycle. The second characteristic is that the negative adjustments of the k parameter are made stricter, with the incorporation of an overfitting parameter aimed at increasing the speed of convergence.

For the first method, which we shall call VL1, the k parameters are marked according to Equation 4, in which γ is the overfitting parameter.

$$k_i^{t+1} = \begin{cases} k_i^{(t)} & \text{if } P^{(t)}(S_i) \le r_{\max} \\ r_{\max}/P^{(t)}(S_i) - \gamma & \text{if } P^{(t)}(S_i) > r_{\max} \end{cases} \tag{4}$$

where $0 \le \gamma < r_{max}/P^{(t)}(A_i)$.

Calculation of the k parameters in the second method is carried out in accordance with Equation 5, in which φ is the overfitting parameter.

$$k_i^{t+1} = \begin{cases} k_i^{(t)} & \text{if } P^{(t)}(A_i) \le r_{\max} \\ k_i^{(t+1)} - P^{(t)}(A_i) + r_{\max} - \varphi & \text{if } P^{(t)}(A_i) > r_{\max} \end{cases} \tag{5}$$

where $0 < \varphi < r_{max}$. We shall call this method VL2.

The VL1 rule obtains adequate results in a smaller number of iterations than the VL2 rule, especially when the γ

parameter has a high value. The VL2 method appears to be less sensitive with regard to the choice of values for the φ parameter (van der Linden, 2003).

The Restricted Method as an Alternative for Increasing Speed of Convergence

In this work we shall propose an alternative method for accelerating the convergence of the k parameters, a system that is valid both for the SH method and the VL1 and VL2 methods. An item-selection rule that allows us to identify, simultaneously, $P(S_i)$ and $P(A_i)$, while also satisfying Equation 1, would serve to provide initial k parameters much closer to the definitive parameters, by comparison with the normal practice, consisting in starting out with k parameters equal to 1. The restricted method (Revuelta & Ponsoda, 1998), which we shall refer to as the R method, would allow us to achieve this. We shall first describe the characteristics of this method, going on to explain how it can increase the speed of convergence of the k parameters for the other methods.

The R method, formulated in Equation 6, marks as administrable items for the jth examinee only those with an exposure rate lower than r_{max}. If the determination of the presentability of an item is done after its selection, we can record both $P(S_i)$ and $P(A_i)$. This rule guarantees that the exposure rate of all the items remains below the target value.

$$k_i^{t+1} = \begin{cases} 1 & \text{if } P^{(1..j)}(A_i) < r_{\max} \\ 0 & \text{if } P^{(1..j)}(A_i) > r_{\max} \end{cases} \tag{6}$$

There are three main differences between the R method and the methods previously described. First, with this method the k parameters are updated on the fly for each new examinee, so no previous simulation phase is needed. Second, the k parameters are not the same for all examinees. Third, the R method accepts only two values for the k parameters, 0 and 1. This means that the method is deterministic, in the sense that, knowing the k parameters of the items for an examinee, his estimated trait level and the items already administered, we know what the next item administered will be. The other methods are probabilistic, since a random experiment mediates between selection and administration.

Given that the k parameters are updated with each examinee, the R method offers several advantages with respect to the SH method or the VL methods: (1) it is irrelevant whether the true distribution of examinees' trait levels is or is not identical to the expected distribution (Chang & Twu, 2001); (2) breaking of item-bank security, which would greatly increase the probability of selection of items with high b parameters, would not also cause an increase in their exposure rate. In either case, the restriction in Equation 1 would still be met. Nevertheless, the R method has received scarce attention from researchers, and as far as we know has only been incorporated in the CAT on knowledge of written English described by Olea, Abad, Ponsoda, and Ximenez (2004).

A clear limitation of the R method, which may explain its low impact, is that both composition and size of the administrable item bank varies from examinee to examinee. Thus, for example, the first examinee has the entire bank available, while for the second examinee the items administered to the first are no longer available.

The R method would make it possible to increase the convergence of the k parameters for the SH method and for the VL methods. For this, it would only be necessary that, apart from registering $P(A)$ for each item, the program were to register also $P(S)$, the proportion of examinees for whom the presentability of each item has been assessed. Given that the method, by definition, guarantees satisfaction of Equation 1, we could achieve initial estimations of the k parameters closer to the final values than when the k parameters start out with values equal to 1. According to the method to be used, the initial k parameters will be as indicated in Equations 7, 8, and 9 (SH, VL1, and VL2 methods, respectively).

$$k_i^{(1)} = \frac{P^{(0)}(A_i)}{P^{(0)}(S_i)} \tag{7}$$

$$k_i^{(1)} = \frac{P^{(0)}(A_i)}{P^{(0)}(S_i)} - \gamma \tag{8}$$

$$k_i^{(1)} = \frac{P^{(0)}(A_i)}{P^{(0)}(S_i)} - \varphi \tag{9}$$

Objectives of the Present Study

An idea implicitly underlying methods that restrict r_{max} is that of satisfying Equation 1 with minimum loss of measurement accuracy. In van der Linden's (2003) proposal, the new methods were not directly assessed with the usual indicators of measurement accuracy. It is possible that the VL1 and VL2 methods over-reduce the exposure rate of the most informative items, so that improvements in exposure control would be at the cost of losses in measurement quality. Testing this possibility was our first objective.

Let us consider, for example, a highly informative item that in the first cycle were selected for all the examinees, giving it an exposure rate of 1. According to rules VL1 and VL2, the k parameter of that item in the second iteration would be equal to (r_{max} – overfitting parameter). Thus, even supposing that the probability of selection for such an item in the second iteration continued to be equal to 1, its probability of administration would be lower than r_{max}, in a proportion depending on the value of the overfitting parameter. Given that the k parameters are only updated for those items with an exposure rate above r_{max}, the k parameter of that item would not be modified and a good item would be under-used.

On the other hand, in van der Linden's (2003) study the value employed for r_{max}, .2, is far removed from the minimum value admissible for r_{max}, [number of items administered/item bank size], which in that study would correspond to .076. The high risk to item bank security involved in diffusing some of its items (Chang, 2004) may make it appropriate to situate r_{max} at values closer to the minimum value possible. It would, therefore, be of interest to examine whether the VL1 and VL2 methods are also valid alternatives to the SH method in these cases. This one can be considered as our second objective.

We are also interested in studying the possible effects of the application of the R method for providing an initial estimation of the *k* parameters. We expect the R method to increase the speed of convergence for the different methods. Studying the effects of the R method was our third objective.

Simulation Studies

Method

Ten item banks were generated, each with 250 items, with parameters *a, b,* and *c* taken at random from distributions N(1.2, .25), N(0, 1), and N(.25, .02), respectively. Length of the CAT was set at 20 items. Examinees' trait level was taken at random from a population N(0, 1). Initial trait level was extracted at random within the interval (–.5, .5). Number of examinees presented at each iteration was 5000. Estimation of trait level was made through the maximum-likelihood method, except when the response pattern was all hits or all errors, in which case the method proposed by Dodd (1990) was used. When all the responses were correct, the estimated trait level was increased by (b_{max} – estimated trait level)/2. If all the responses were wrong, estimated trait level was reduced by (estimated trait level – b_{min})/2.

For a bank of this size and with this number of items administered, the minimum possible r_{max} is .08. We worked with two values of r_{max}, one a strict value of .1, and the other less demanding, of .15. Value of the parameters γ and φ was set at $2/3r_{max}$. This value is within the range of values suggested by van der Linden (2003). As in van der Linden's (2003) study, control of exposure was applied to the entire population of examinees. It is assumed that the results for a control of exposure conditioned to ability levels would provide similar patterns.

For each method we simulated 20 iterations plus an initial cycle (cycle 0) to start off the parameters. All the rules (SH, VL1, and VL2) were simulated under two conditions: using the R method for initial estimation of the *k* parameters (R conditions) and without using the R method (NOR conditions). Cycle 0 corresponds, for the R methods, to the iteration in which the method employed as restriction of r_{max} is only R. For the NOR methods, cycle 0 is applied with initial values of the *k* parameters equal to 1 (no exposure control).

The results were analyzed according to 6 dependent variables. Three of these are identical to those used by van der Linden (2003): (a) maximum exposure rate; (b) proportion of items above r_{max}; and (c) mean exposure rate of the items over r_{max}. To these were added: (d) overlap rate, approximated according to Equation 10, as a measure of item bank security; (e) RMSE, indicative of measurement accuracy, the value calculated according to Equation 11; and (f) mean value of the *a* parameter of the items administered. If the variations on the SH method maintain the same item-selection logic as the original method, this value should be similar to that found with the original method.

The overlap rate was

$$\hat{T} = \frac{n}{q} S^2_{P(A)} + \frac{q}{n} \tag{10}$$

where
$\hat{T}$ is the estimation of the populational overlap rate (Chen et al., 2003);
n is the item-bank size;
q is the number of items to be administered; and
$S^2_{P(A)}$ is the variance of the exposure rates of the items.

The RMSE was

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{e} (\hat{\theta}_j - \theta_j)^2}{e}} \tag{11}$$

where
$\hat{\theta}_j$ is the estimation of the trait level of the *j*th examinee;
θ_j is the trait level of the *j*th examinee; and
e is the number of examinees.

Results

Figure 1 shows the results for r_{max} = .15 with respect to maximum exposure rate, proportion of items with rate above r_{max} and mean rate of exposition of the items with rate over r_{max}. As in van der Linden (2003), the fit for each one of these criteria was better for the rules VL1 and VL2 than for the SH method. Speed of convergence of the VL2 method is the lowest of the three methods. Importantly, the use of the R method for initial calculation of the *k* parameters markedly increases speed of convergence of each one of the methods. The R and NOR methods eventually converge to identical values for the variables studied, even though the NOR methods stabilize later. For example, for the R-SH method the maximum exposure rate and the mean of the items with rate higher than r_{max} stabilize around the second or third cycle, while for the NOR-SH method this

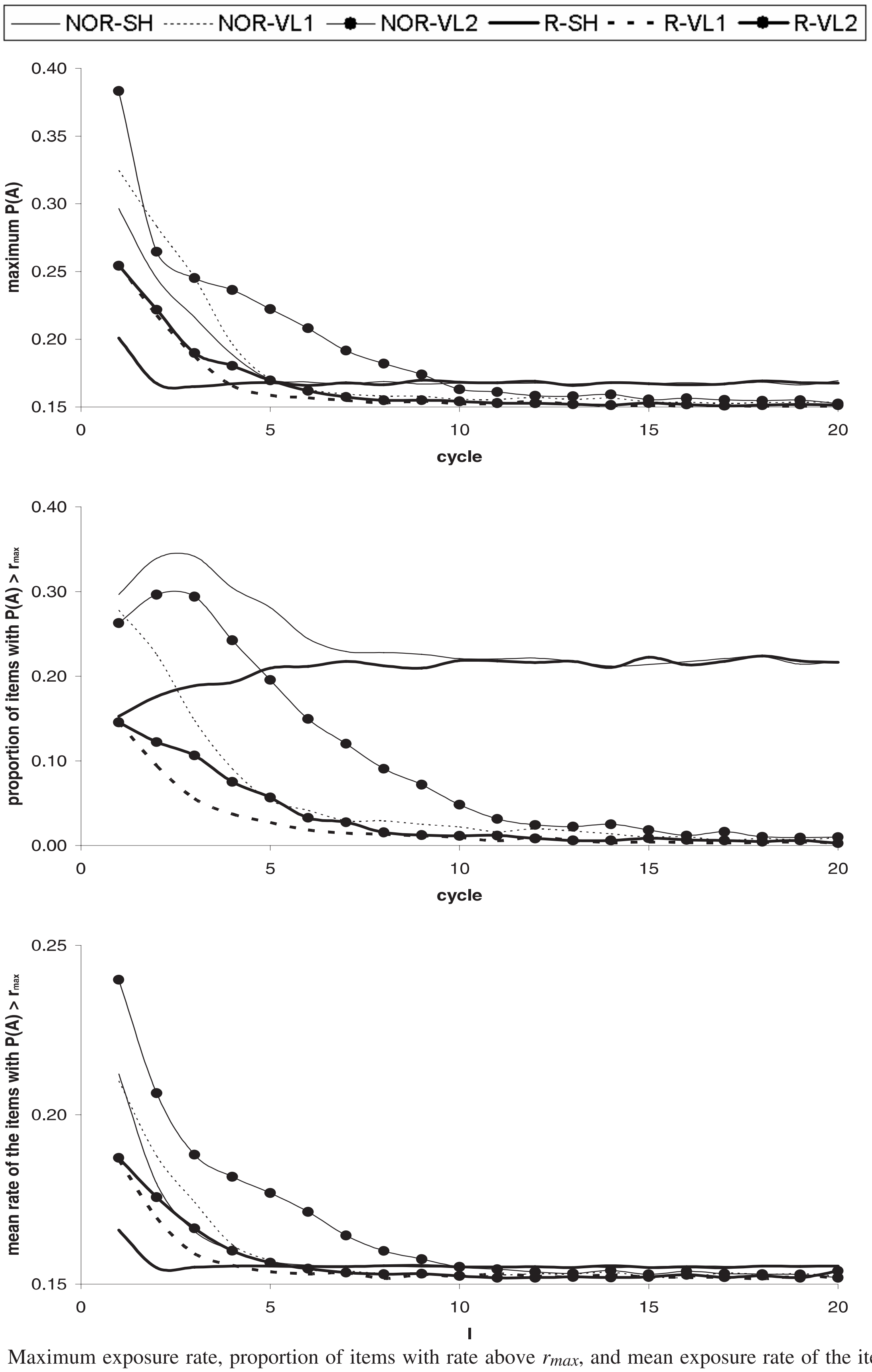

Figure 1. Maximum exposure rate, proportion of items with rate above r_{max}, and mean exposure rate of the items with rate above r_{max} for r_{max} = .15.

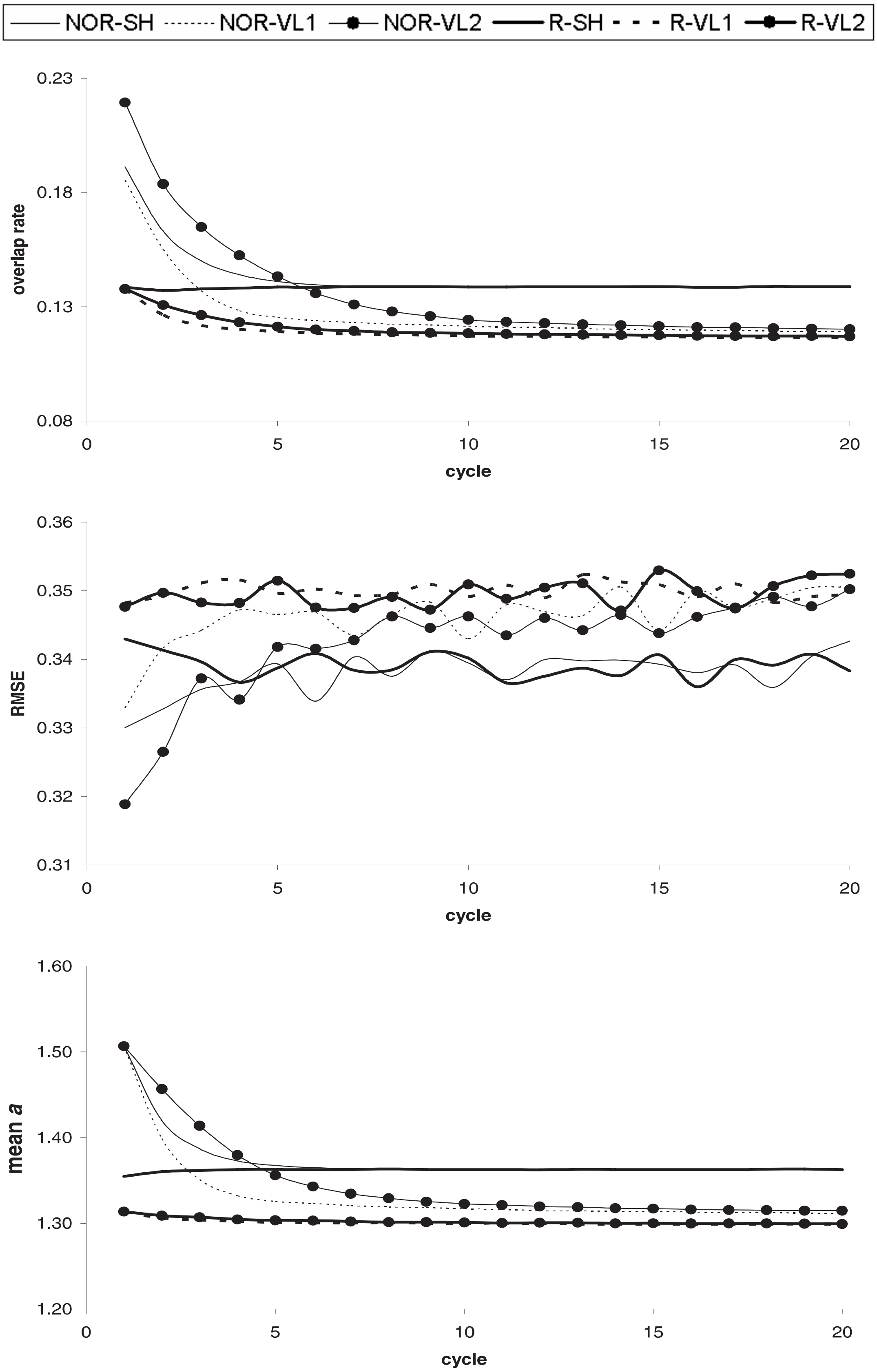

Figure 2. Overlap, RMSE, and average value of the *a* parameter for $r_{max} = .15$.

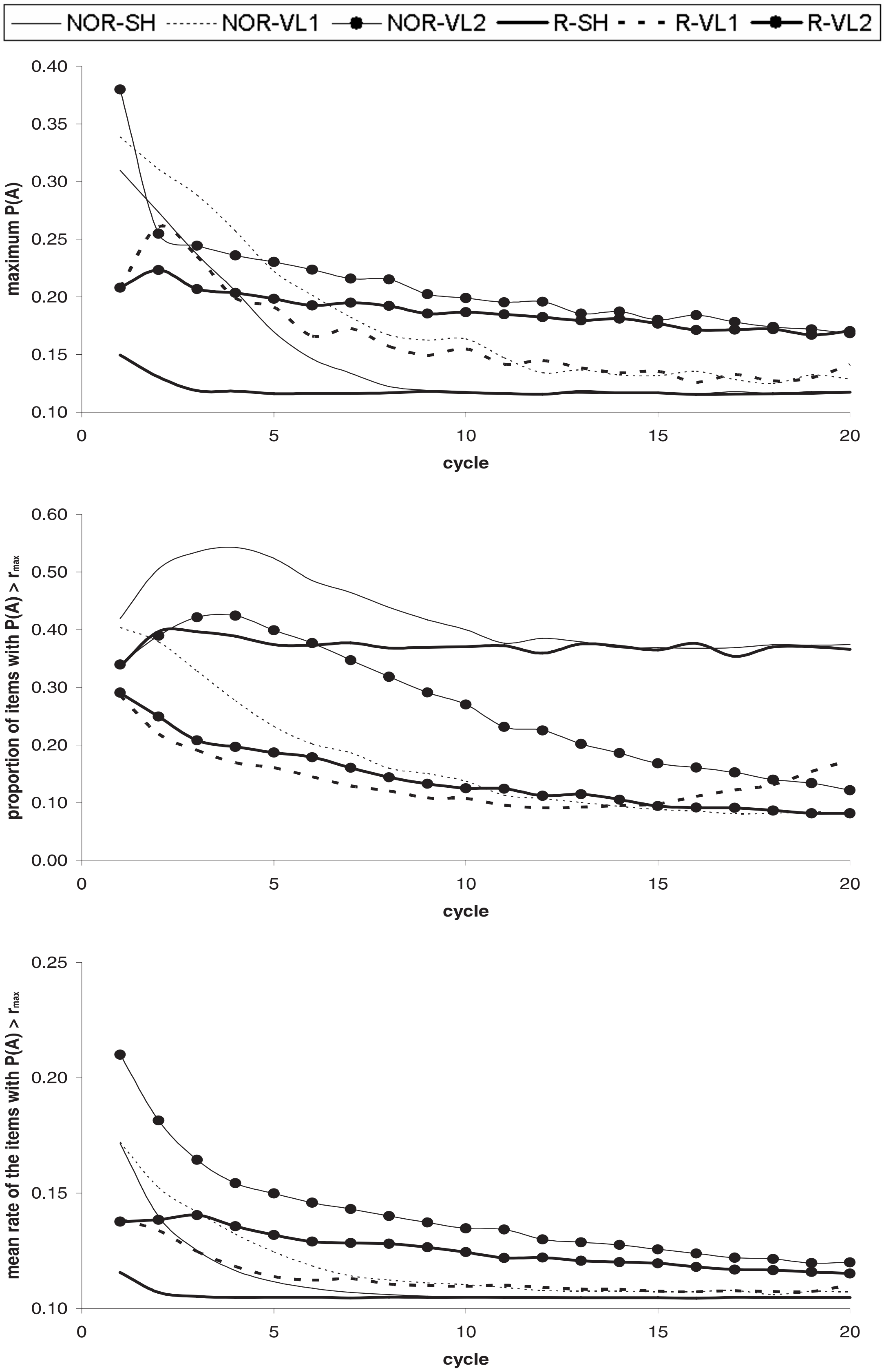

Figure 3. Maximum exposure rate, proportion of items with rate above r_{max}, and mean exposure rate of the items with rate above r_{max} for $r_{max} = .1$.

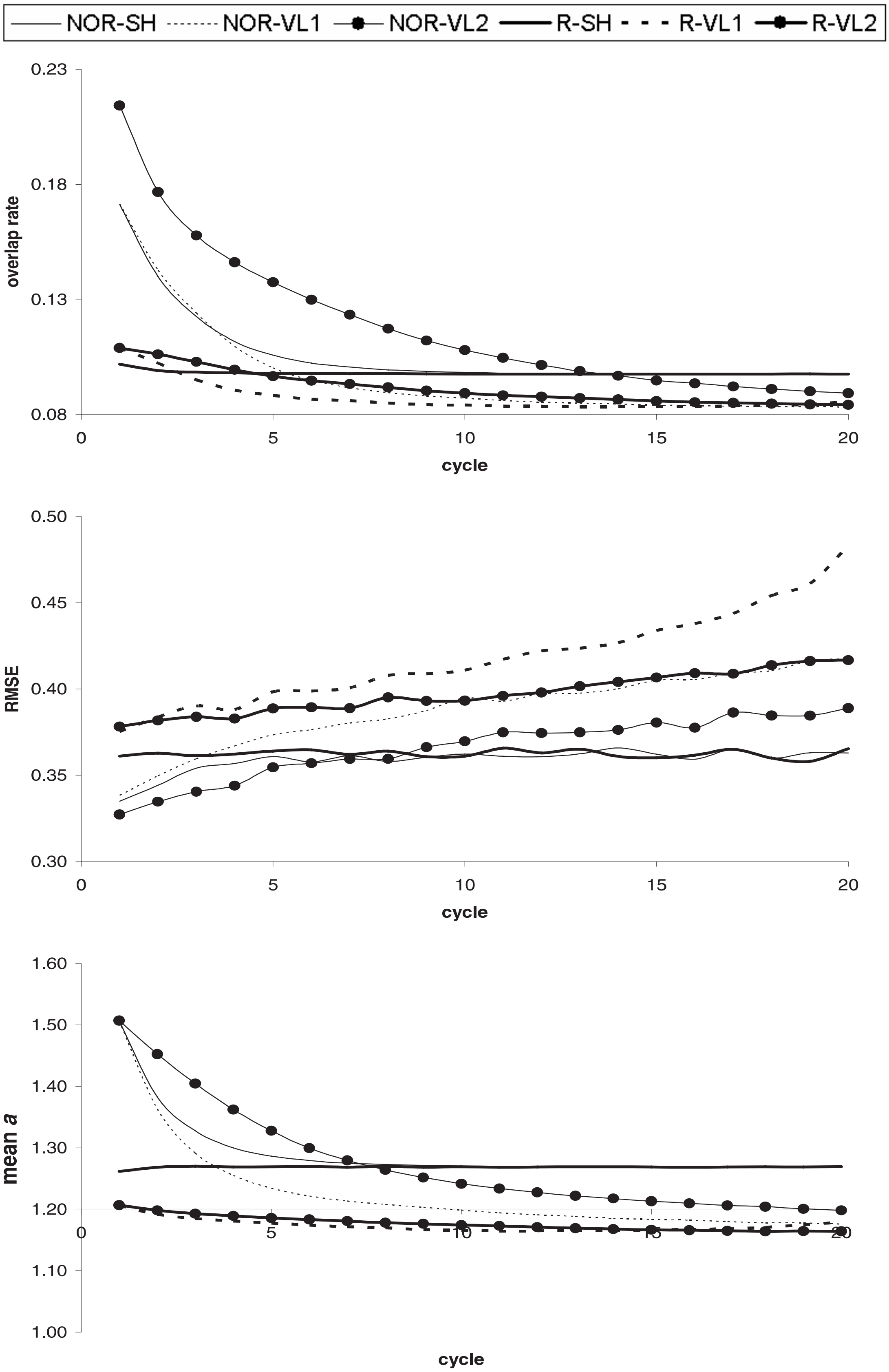

Figure 4. Overlap, RMSE, and average value of the *a* parameter for r_{max} = .1.

does not occur until the eighth iteration. For the VL1 and VL2 methods the equalization in results between R and NOR is later.

Figure 2 shows the results in overlap rate, RMSE, and mean of the *a* parameter for r_{max} = .15. As expected, given the better fit obtained with the VL rules, for these methods the overlap rate is lower than that obtained with the SH method. We confirmed the hypothesis proposed in the description of the methods: the VL1 and VL2 methods involve losses in measurement accuracy, by comparison with SH. These differences are small but consistent. Likewise, we found that the items administered with the VL rules tend to have a lower value in the *a* parameter than when the rule used is the SH.

The differences in both the RMSE and the average *a* value between the SH methods and the VL methods have two possible explanations. One would be the greater capacity of the VL methods for satisfying the restriction raised in Equation 1, as can be seen in Figure 1. On restricting the availability of certain items below r_{max}, the measurement error increases and the mean discrimination parameter of the items administered decreases. The other explanation is that there has been over-restriction in the exposure rate of items with high probability of selection, situating the exposure rate of these markedly below r_{max}.

As we saw in Figure 1, the stabilization of the variables shown in Figure 2 is quicker for the R methods. The SH method is that which shows a more rapid and more complete convergence between the R method and the NOR method. The R-VL methods, compared to the NOR-VL methods, provide a higher RMSE and administer items with lower *a* parameter. This appears to indicate that the VL methods, when employed with the R system, are more susceptible to over-restriction of items with high probability of selection.

The results undergo important changes when r_{max} is made stricter, to equal .1. Figure 3 shows the data for this condition for the variables maximum exposure rate, proportion of items with rate over r_{max} and mean rate of the items with rate over r_{max}. When the maximum rate approaches the minimum that can be imposed, the VL methods present problems of convergence. Given the number of iterations simulated, we do not know whether this convergence is simply slow, or whether stabilization is impossible to achieve. In contrast to what occurred with a less strict r_{max}, the SH method seems to be more effective in bringing the maximum rate found close to the target value, and with this method lower values of mean rate of items over r_{max} can be achieved. With the SH method, the proportion of items surpassing r_{max} is markedly higher than that found with the VL methods. The R methods begin with better results than the NOR methods for each one of these criteria. As the number of cycles increases, the two methods eventually converge. The convergence between the two seems to be slower the stricter the control on exposure.

Figure 4 shows the results for overlap rate, RMSE, and average value of the *a* parameter. Once again, the VL methods show an overlap rate lower than that achieved with the SH methods. Stabilization of the overlap rate is slower for the VL rules than for the SH method. Especially important are the results of the RMSE and mean *a* parameter. In the 20 iterations simulated, the RMSE for the VL rules does not become stabilized, increasing with each successive iteration. With the VL methods, with each new iteration, the items administered have lower discriminative capacity. Surprisingly, for three of the VL methods the average value of the *a* parameter actually reaches a situation of being below the mean value of this parameter in the item bank. The NOR-VL2 method does not show values below this mean, presumably because the simulations ceased at 20 iterations. The R-VL methods offer poorer measurement accuracy and a mean on the discrimination parameter of the items administered lower than the NOR-VL methods.

As we understand it, these anomalies in the process of adjustment for the VL methods are due to the overfitting parameter. The introduction of these parameters seems to result in a penalization of the items with high discrimination parameter, forcing them toward exposure rates below the rate set as a limit. As no increase of the values of the *k* parameters is allowed, this problem is maintained. Importantly, in these conditions the SH method functions normally, attaining convergence, albeit with some items with exposure rates slightly higher than r_{max}.

Discussion

The purpose of the present paper was to assess the functioning of different methods proposed for maintaining the item exposure rate below a maximum level. Likewise, we proposed a method, based on the R method (Revuelta & Ponsoda, 1998) for accelerating stabilization in the results of the different methods. To this end we studied the methods in two different conditions, one with strict r_{max} and another with more relaxed r_{max}.

When r_{max} is equal to .15, the results of the NOR conditions found replicate those of the study by van der Linden (2003). The VL methods improve convergence of the rates below the desired limit. Likewise, the overlap rate is lower for the VL rules. Even so, there is cost in the form of a slight increase in RMSE. For the three methods studied, the rules with incorporation of the R rule offer more rapid stabilization of the results than those that do not employ the R rule.

The pattern of results when r_{max} = .1 is substantially different. In this case, the SH method seems to be that which functions best, both because it is that which comes closest to satisfying Equation 1 and because it offers the lowest RMSE. When r_{max} is strict, the VL methods are unable to stabilize their results, and present anomalies in their functioning, such as situating the mean of the *a* parameter of the items administered below the mean of this parameter in the item bank. In this condition, the SH method continues to benefit from the incorporation of the R method.

The overfitting parameter present in the VL methods appears to have unexpected negative effects on the original proposal (van der Linden, 2003) in situations of very strict control over exposure (that is, when r_{max} is close to the minimum possible value). An alternative solution to this problem would involve using a wide variety of values for the overfitting parameters and, after the simulations, choosing the most appropriate. However, in doing so we would lose one of the benefits of the VL methods, which is the reduction in computation time necessary for calculating the k parameters.

Given the possible adverse effects of the VL rules, we recommend particular care when using them. For the time being, it would seem advisable to continue employing the SH method, incorporating the R method for accelerating the process.

References

Barrada, J.R., Olea, J., & Ponsoda, V. (2005). *Item selection rules in computerized adaptive testing: Accuracy and security*. Manuscript submitted for publication.

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

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

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

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.

Chang, S.W., & Twu, B.Y. (2001, April). *Effects of changes in the examinees' ability distribution on the exposure control methods in CAT.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

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.

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.

Leung, C.K., Chang, H.H., & Hau, K.T. (2002). Item selection in computerized adaptive testing: Improving the a-stratified design with the Sympson-Hetter algorithm. *Applied Psychological Measurement, 26*, 376–392.

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.

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.

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., & 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., & Veldkamp, B.P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics, 29*, 273–291.

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

Juan Ramon Barrada

Dpto. de Psicologia Social y Metodologia
Facultad de Psicologia
Universidad Autonoma de Madrid
28049 Madrid
Spain
Tel. +34 91 497-87-50
Fax +34 91 497-52-15
E-mail juanra.barrada@uam.es

