

Rotating Item Banks versus Restriction of Maximum Exposure Rates in Computerized Adaptive Testing

Juan Ramón Barrada¹, Julio Olea², and Francisco José Abad²

¹Universidad Autónoma de Barcelona (Spain)

²Universidad Autónoma de Madrid (Spain)

If examinees were to know, beforehand, part of the content of a computerized adaptive test, their estimated trait levels would then have a marked positive bias. One of the strategies to avoid this consists of dividing a large item bank into several sub-banks and rotating the sub-bank employed (Ariel, Veldkamp & van der Linden, 2004). This strategy permits substantial improvements in exposure control at little cost to measurement accuracy. However, we do not know whether this option provides better results than using the master bank with greater restriction in the maximum exposure rates (Simpson & Hetter, 1985). In order to investigate this issue, we worked with several simulated banks of 2100 items, comparing them, for RMSE and overlap rate, with the same banks divided in two, three... up to seven sub-banks. By means of extensive manipulation of the maximum exposure rate in each bank, we found that the option of rotating banks slightly outperformed the option of restricting maximum exposure rate of the master bank by means of the Simpson-Hetter method.

Keywords: item bank security, overlap rate, accuracy, bank rotation, Simpson-Hetter method

Si los examinados conocieran de antemano una parte del contenido de un test adaptativo informatizado, sus niveles estimados de rasgo tendrían un marcado sesgo positivo. Una de las estrategias para evitar esto consiste en dividir un gran banco de ítems en varios sub-bancos y rotar el sub-banco empleado (Ariel, Veldkamp & van der Linden, 2004). Esta estrategia permite mejoras sustanciales en el control de la exposición con poca merma de la precisión de la medida. Sin embargo, no sabemos si esta opción proporciona mejores resultados que el uso del banco maestro con más restricción en la tasa máxima de exposición (Simpson & Hetter, 1985). Para investigar este problema, trabajamos con varios bancos simulados de 2100 ítems, comparándolos, en RMSE y en tasa de solapamiento, con los mismos bancos divididos en dos, tres... hasta siete sub-bancos. Mediante manipulación extensa de la tasa máxima de exposición en cada banco, encontramos que la opción de rotar los bancos ofrecía resultados ligeramente mejores que la opción de restringir la tasa máxima de exposición del banco maestro mediante el método Simpson-Hetter.

Palabras clave: seguridad del banco de ítems, tasa de solapamiento, precisión, rotación del banco, método Simpson-Hetter

This research was partly supported by a grant from the Spanish Ministerio de Educación y Ciencia (project SEJ2006-08313/PSIC).

Correspondence concerning this article should be addressed to Juan Ramón Barrada, Facultad de Psicología, Universidad Autónoma de Barcelona, 08193 Bellaterra (Spain). Phone: 00 34 93 581 32 63. E-mail: juanramon.barrada@uab.es.

How to cite the authors of this article: Barrada, J.R., Olea, J. and Abad, F.J.

When the application of tests has important consequences for those being assessed and for institutions, examiners and examinees generally have clearly different objectives (Wainer, 2000). Whilst the former seek an accurate and efficient evaluation of the examinees' trait level, some examinees try to obtain the most positive assessment possible, regardless of whether this corresponds to their true level. To this end, examinees may make use of various illegitimate means to increase their score (Davey & Nering, 2002), one of the most productive being to find out some of the questions beforehand.

This risk is clearly present in assessment by means of Computerized Adaptive Testing (CAT). In CAT, the same items are maintained operative for a relatively long period, which may facilitate "leakage" of some of the item bank content (Chang, 2004). There are various ways to reduce this risk: (a) seeking secure item-selection rules (Barrada, Olea, Ponsoda, & Abad, in press; Chang & Ying, 1999); (b) generating indicators for the detection of leaked items (Lu & Hambleton, 2004; Veerkamp & Glas, 2000); (c) developing trait-level estimation methods that incorporate the possible effect of previous knowledge of items (Segall, 2004). All such actions, in addition to those of the withdrawal of items from the bank and the incorporation of new items, should be considered simultaneously in the maintenance of an operational CAT (Wise & Kingsbury, 2000).

The higher the proportion of identical questions that different examinees receive, the greater the risk will be. One of the variables usually used to measure this risk is overlap rate. Overlap rate is the mean proportion of items shared by two examinees (Way, 1998). Reducing the overlap rate between examinees therefore increases the security of the bank. Chen, Ankenmann and Spray (2003) have demonstrated that homogenizing the item exposure rates leads to a reduction in the overlap rate. One possibility of achieving this is the construction, based on a large master bank, of different sub-banks of items, which would be used in a rotating manner (Mills & Steffen, 2000; Stocking & Swanson, 1998). In this way, maximum exposure rate can be reduced, since, if a bank is divided into q sub-banks, the maximum possible exposure rate will no longer be 1, but rather $1/q$.

Ariel, Veldkamp and van der Linden (2004) compared various ways of distributing the items of the master bank in different sub-banks. Their results showed that, regardless of the method used to construct them, the use of rotating banks improves the security of the item bank, albeit with slight losses in measurement accuracy. They found, therefore, the trade-off between accuracy and security to which other authors have referred (Way, 1998).

Another alternative for homogenizing the exposure rates of the items is direct control of the maximum exposure rate. Different methods have been proposed for this (Barrada, Veldkamp, & Olea, in press; Revuelta &

Ponsoda, 1998; Sympson & Hetter, 1985; van der Linden & Veldkamp, 2004). In all these methods, a maximum exposure rate (r_{max}) is fixed, and, by means of different approaches, it is sought to keep the exposure rate of all items below it. The restrictions on r_{max} using these methods leads, simultaneously, to reductions in the overlap rate and losses of measurement accuracy, the above-mentioned trade-off between accuracy and security being found again.

Another option, not yet studied, is to combine the two proposals: to impose restrictions on r_{max} in the rotating banks. With the data currently available, it is not clear which option is the most appropriate. The objective of the present work is to present a comparison between the different alternatives. To this end, we first describe the two aforementioned approaches to exposure control, as they were applied in this study. Subsequently, we describe the research method employed. Finally, we present the results and their discussion.

Two Approaches to Exposure Control

We do not pretend to be exhaustive in our description of the possible alternatives for controlling security in CATs. These are dealt with more comprehensively in the reviews by Davey and Nering (2004) and Way (1998). We concentrate on the procedures applied in this study.

Construction of Rotating Item Banks

Rotating banks can be designed with or without overlap. Non-overlapping banks are those in which the different sub-banks do not share any items. In overlapping banks, some items form part of several sub-banks: the lower the exposure rate of an item in the master bank, the more sub-banks it will be assigned to. While banks without overlapping succeed in reducing over-exposure of items, they have scarcely any effect on under-exposure. Banks with overlapping aim to improve these two problematic situations at the same time.

The different sub-banks constructed must show similar distributions of parameters and content, so as to offer uniform quality of measurement among examinees. Ariel et al. (2004) have proposed various methods to guarantee this, all based on linear programming techniques (van der Linden, 2005). In the present work we have studied only the functioning of non-overlapping banks. Following Ariel et al., their construction consists of two phases. In the first one, the n items of the master bank are assigned to n/q interim sets, where q is the number of sub-banks to be constructed, such that the q items making up the different interim sets are maximally similar to one another.

The value δ_{ij} is used as a measure of the difference between items i and j , such that

$$\delta_{ij} = |a_i - a_j| + w|b_i - b_j| \quad (1)$$

where a is the discrimination parameter, b is the localization parameter and w is a parameter that can be used to correct the differences of scale between the two parameters.

The objective would be to minimize the value δ_{ij} for all possible combinations of items belonging to the same interim set. With this aim, the variable x_{ij} is created, which adopts a value of 1 when i and j belong to the same interim set and 0 otherwise, where i and j belong to $[1..n]$ and $i \neq j$. This problem can be seen as one of linear programming, with target function

$$\min \sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij} x_{ij} \quad (2)$$

subject to the restriction that the items are assigned to a single interim set.

In the second phase, the q items making up the different interim sets are distributed among the q sub-banks of items, such that the information function of the different sub-banks is maximally similar among them. A measure of the achievement of this goal in a fixed trait level (θ_u) could be the sum of the differences in information for the different sub-banks:

$$\gamma_{u,st} = \left| \sum_{i=1}^n I_i(\theta_u) y_{is} - \sum_{j=1}^n I_j(\theta_u) y_{jt} \right| \quad (3)$$

where i and j are two items belonging to the bank and y_{is} and y_{jt} are decision variables, with a value of 1 if items i or j are assigned to sub-banks s or t , respectively.

The objective would be to minimize the difference in the information functions for all possible comparisons $2 * 2$ of all the sub-banks for several points of ability that suitably represent the continuum of the trait level distribution. The target function to evaluate for this would be:

$$\min \sum_{u=1}^z \sum_{s=1}^{q-1} \sum_{t=s+1}^q \gamma_{u,st} \quad (4)$$

subject to

$$\sum_{i \in Q_r} y_{is} = 1, \forall s \quad (5)$$

$$\sum_s y_{is} = 1, \forall i \quad (6)$$

$$y_{is} \in \{0,1\} \quad (7)$$

where z is the number of trait levels for which the information functions will be compared, q is the number of sub-banks generated and Q_r is the r^{th} interim set.

The restriction in Equation 5 forces the items in the same interim set to be assigned to different sub-banks. Equation 6 indicates that items can only be assigned once, whilst Equation 7 marks the possible values for the decision variable.

Thus, the master bank is divided into q sub-banks, guaranteeing that these are as similar as possible. Ariel et al. (2004) describe how this general model can be extended for the case of overlapping banks and so as to incorporate additional restrictions. The efficiency of this method for constructing homogeneous sub-banks can be checked in the article by Ariel et al.

Direct Control of the Maximum Exposure Rate

The Sympon-Hetter method (SH – Sympon & Hetter, 1985) 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). This method seeks to fix the exposure rate of any item, $P(A_i)$, below a maximum exposure rate (r_{max}).

$$P(A_i) \leq r_{max}, \forall i \quad (8)$$

In the SH method, the choice of an item does not necessarily lead to its presentation: the probability of administration of an item after being selected [$P(A_i | S_i)$] is the way in which fits equal or below r_{max} .

$$P(A_i) = P(A_i | S_i)P(S_i) \quad (9)$$

Through a series of simulation cycles, it is sought to establish the parameters $P(A_i | S_i)$, such that Equation 8 is satisfied. The parameters for the cycle $c+1$ are calculated as follows:

$$P^{(c+1)}(A_i | S_i) = \begin{cases} 1 & \text{if } P^c(S_i) \leq r_{max} \\ r_{max}/P^c(S_i) & \text{if } P^c(S_i) > r_{max} \end{cases} \quad (10)$$

These simulation cycles are repeated until the empirical maximum exposure rate is stabilized, normally slightly above r_{max} (Barrada, Olea & Ponsoda, 2007; van der Linden, 2003). Although some alternatives to the SH method have been developed (Revuelta & Ponsoda, 1998; van der Linden & Veldkamp, 2004), it continues to be the method most commonly used (van der Linden, 2003).

Simulation Study

Method

It is common in operative item banks that the a and b parameters are positively correlated. Because of this,

two kinds of master banks of 2100 items were generated, both with parameters a , b and c randomly extracted from distributions $N(1.2, .25)$, $N(0, 1)$ and $N(.25, .02)$, respectively. While the first kind had uncorrelated a and b parameters, in the second kind the correlation between these parameters in the population was equal to .5. The bank size chosen is similar to the one used by Ariel et al. (2004) in their study, based on an old bank of 2131 items from the Law School Admission Test. Length of the CAT was fixed at 25 items. We generated 20 different banks, 10 of each kind.

Two different distributions of the examinees' trait level were used. In the first one, trait levels were taken at random from a population $N(0, 1)$. This implies that the distribution of the b parameters of the bank and the trait levels of the examinees had the same mean and standard deviation. A second condition, where the trait levels were randomly extracted from a population $N(-.4, 1)$ was also tested, with the aim of checking whether the results varied depending with the match/mismatch of both distributions. Number of examinees presented was 5000.

Initial trait level of examinees was extracted at random within the interval $(-.5, .5)$. Trait level estimation was made by means of the maximum-likelihood method, except when the pattern of responses was constant, all correct or all incorrect, at which point the method proposed by Dodd (1990) was used: 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 any case, the estimated trait levels had to be within the interval $[-4, 4]$. The item selected at each point was the one that maximized the Fisher information function for the estimated trait level.

The master bank was divided in two, three... up to seven rotating sub-banks. With seven sub-banks, the size of each was 300 items, the minimum bank size admissible for a 25-item CAT, according to the recommendations of Stocking (1994). For the construction of the sub-banks we used Equations 1 to 7. The value of w in Equation 1 was fixed, as in Chang and van der Linden (2003), at:

$$w = \frac{a_{max} - a_{min}}{b_{max} - b_{min}} \tag{11}$$

The values of the information function of the different sub-banks were compared at 5 points, from $\theta = -3$ to $\theta = 3$, in steps of 1.5. Taking into account that Ariel et al. (2004) used just three points $(-1, 0$ and $1)$, we consider 5 points to be sufficient.

In order to assess the performance of the different methods, we used two dependent variables: (a) Root mean squared error (RMSE) relative to measurement accuracy, computed through Equation 12; and (b) overlap rate relative to bank security, the calculation of which was made according to Equation 13.

$$RMSE = \sqrt{\frac{\sum_{g=1}^e (\hat{\theta}_g - \theta_g)^2}{e}} \tag{12}$$

where e is the number of examinees, $\hat{\theta}_g$ is the estimated trait level for examinee g and θ_g is the true trait level.

$$\hat{T} = \frac{m}{n} + \frac{n}{m} S_{P(A)}^2 \tag{13}$$

where m is the number of items to be administered, n the item bank size and $S_{P(A)}^2$ is the variance of the exposure rates of the items (Chen et al., 2003).

With the SH method there is no way of knowing, in advance, the maximum exposure rate that must be fixed in a given bank in order to obtain a pre-fixed overlap rate (Chen & Lei, 2005; Chen, Lei, & Liao, in press). Therefore, it was impossible to know which values of r_{max} we had to apply in the master bank so as to achieve an overlap rate equivalent to that obtained in the different sub-banks, which would allow a comparison in the RMSE without having to consider the overlap rate as a co-variable.

A systematic manipulation of r_{max} allowed us to obtain the functions relating r_{max} to RMSE and overlap rate. We were thus able to obtain the function that relates RMSE and overlap rate, which allows comparison between the master bank with restriction in maximum rate by means of SH and the rotating banks method (see below). For each of the possible r_{max} , the $P(A_i | S_i)$ parameters of the SH method applied were those obtained at the 25th cycle.

Pilot studies had shown that the effects of the change in r_{max} in the RMSE and the overlap rate were stronger for low values than for high values. When q sub-banks are rotated (each examinee receives items extracted from a sub-bank randomly selected before test administration), it is impossible that any item has an exposure rate over $1/q$. Because of this, it makes no sense to set values of r_{max} below $1/q$. Ten values for r_{max} were fixed taking these two points into account, by means of Equation 14. These 10 values varied according to the number of sub-banks in which the master bank was divided.

$$r_d = \frac{m}{n} + \frac{(q^{-1} - m/n) \sum_{h=1}^d (h-1)^2}{\sum_{h=1}^{10} (h-1)^2} \tag{14}$$

where h is a dummy variable, just used for calculations, and d is used for defining the position in the 10 different r_{max} values employed (r_1 , the smaller; r_{10} , the maximum).

Thus, the range between the minimum possible value of r_{max} ($r_1 = m/n$) and the maximum value, given the number of sub-banks used ($r_{10} = q^{-1}$) was divided into intervals of increasing size, as can be seen in Figure 1.

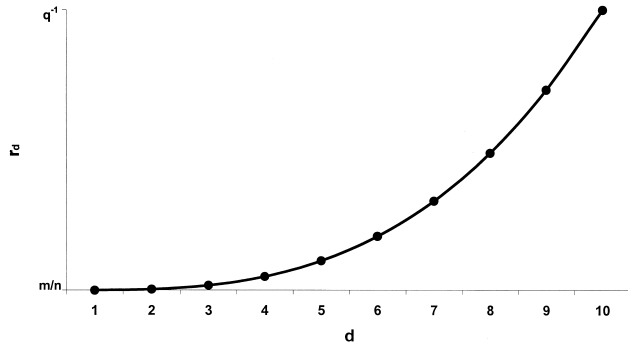


Figure 1. Function relating the indicator variable d and the corresponding maximum exposure rate (r_d).

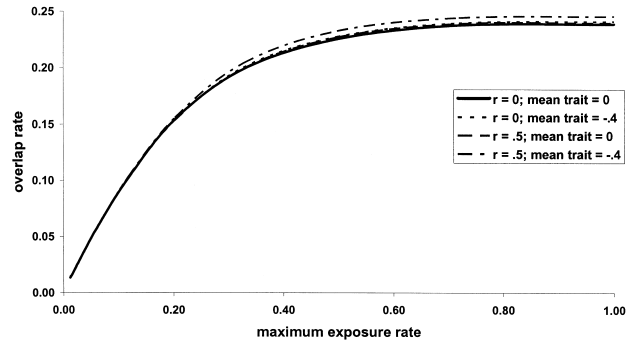


Figure 2. Relation, for the master item bank, between maximum exposure rate and overlap rate for the two item banks and the two examinees' distributions.

Manipulating the variable r_{max} in this way, we were able to obtain q tables of results with the structure shown in Table 1. The first row, d , just indicates which of the 10 levels of the independent variable r_{max} we are referring to. The second row, r_{max} , adopts the values defined with Equation 14. The third row, overlap rate, is a dependent variable. As r_1 is the maximum possible restriction of r_{max} , \hat{T}_1 should be near the minimum overlap rate, m/n . The fourth row, RMSE, is also a dependent variable.

With these data, we managed to obtain scatterplots of r_{max} and overlap, of r_{max} and RMSE and, more importantly, of overlap rate with RMSE. This final chart is the one that we will use to compare the option of using a master bank with restrictions of r_{max} , the option of rotating sub-banks and the option of rotating sub-banks while restricting on them r_{max} .

Results

We will start showing the results just for the master bank, illustrating the different functions that can be obtained by means of our simulations. Figures 2, 3 and 4 show these functions for the four conditions simulated (two kinds of item banks * two distributions of examinees' trait level).

In Figure 2 we can see the relation between r_{max} and overlap rate. As expected, restrictions on r_{max} lead to reductions in overlap rate. Speed of overlap rate reduction is greater the lower the values of r_{max} . In fact, changing r_{max} from .6 to 1 had almost no impact in the overlap rate.

Differences among plots depending on the correlation between parameters or main trait level of the simulees are very small. The condition with greater overlap rate is when item parameters are correlated and the mean trait level is equal to $-.4$.

Figure 3 shows the relation between r_{max} and RMSE. With banks of correlated a and b parameters the measurement error is higher, as the mean discrimination of the items near the examinees' mean trait level is lower. Measurement error is also higher when the mean trait level of the examinees is displaced from the mean b of the items, as there are fewer items available near $-.4$. Speed of RMSE increments is greater the lower the values of r_{max} . Great restrictions on r_{max} can be imposed with negligible effects on RMSE. Comparing Figure 2 and 3 we see, for example, how, for r_{max} equal to .2 the deterioration in RMSE, by comparison with the condition without restriction ($r_{max} = 1$), is almost zero, while a great improvement in the overlap rate can be obtained.

Figure 4 shows the relation between overlap rate and RMSE. As was to be expected, there is a trade-off between RMSE and overlap rate, though, as was seen in the two previous graphs, we found that substantial improvements in bank security can be achieved with hardly any loss of measurement accuracy. In accordance with the results described above, the same overlap rate leads to greater RMSE in the case of correlated parameters and in the case of mismatch between b parameters distribution and examinees' trait level distribution.

Table 1

Relation between the Indicator Variable d and the Corresponding r_{max} , Overlap Rate and RMSE

d	r_{max}	Overlap rate	RMSE
1	$r_1=m/n$	\hat{T}_1	$RMSE_1$
2	r_2	\hat{T}_2	$RMSE_2$
...
10	$r_{10}=q^{-1}$	\hat{T}_{10}	$RMSE_{10}$

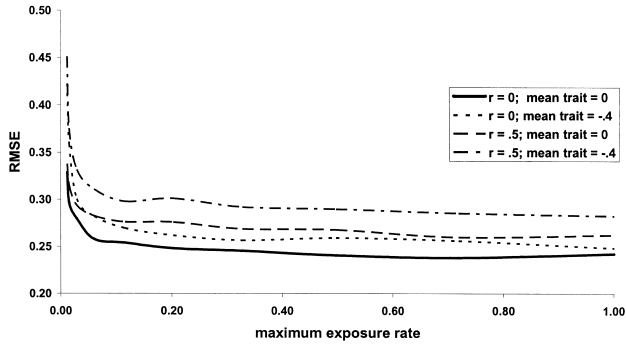


Figure 3. Relation, for the master item bank, between maximum exposure rate and RMSE for the two item banks and the two examinees' distributions.

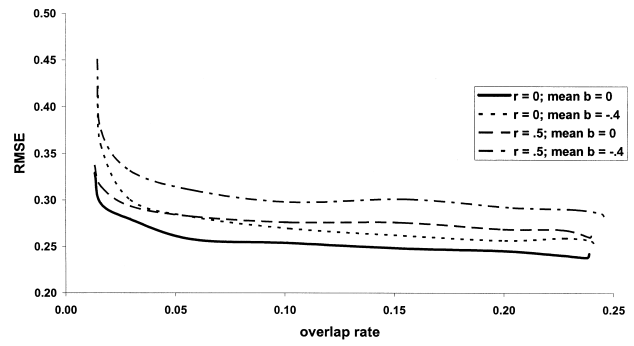
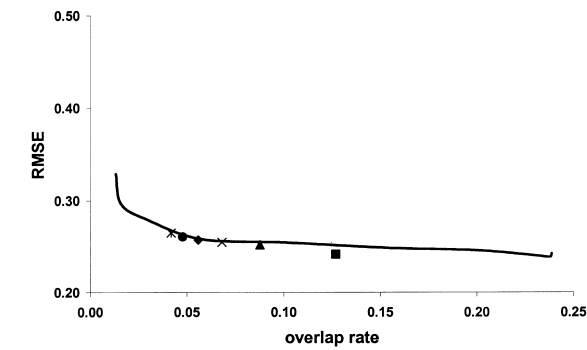


Figure 4. Relation, for the master item bank, between overlap rate and RMSE for the two item banks and the two examinees' distributions.

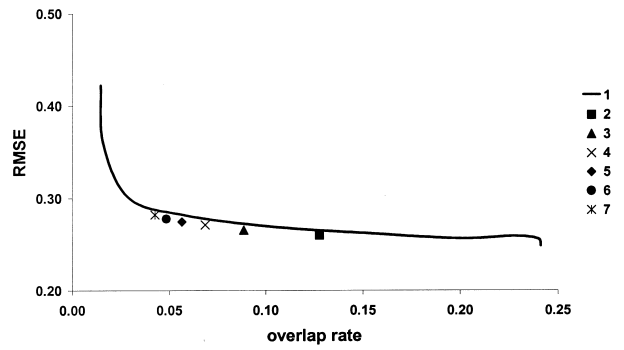
Once we have described the way how the function relating overlap rate and RME can be obtained, we present, in Figures 5 and 6, these results for the master bank and two, three... up to seven rotating banks constructed from this master bank. In Figure 5 we show the condition where rotating item banks have not been combined with the SH method. It can be clearly seen that the higher the number of rotating banks, the lower the overlap rate. For instance, seven rotating banks have an overlap rate that is at about one fifth the overlap rate of the master bank, when no restriction on r_{max} is applied. Increasing the number of rotating item banks

increases the RMSE. Passing from no rotating bank (the master pool) to seven sub-banks lead to an increment in the RMSE of .02. As can be seen, the results of the rotating item banks are below the line described by the master pool when combined with the SH method. This implies that, given the same values of overlap rate, a lower RMSE will be obtained when using rotating item banks. It should also be noted that the differences are, always, very small.

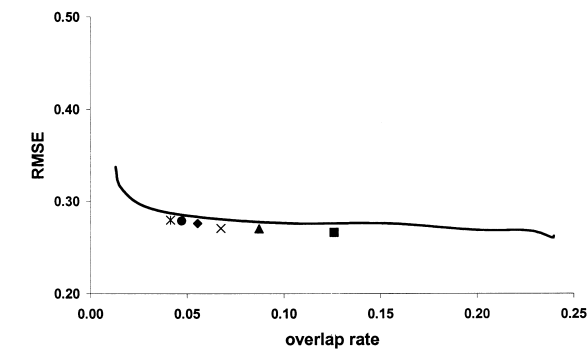
Basically, the same results are obtained when combining the rotating item bank strategy with the SH method. This condition is shown in Figure 6. Again, the option of rotating



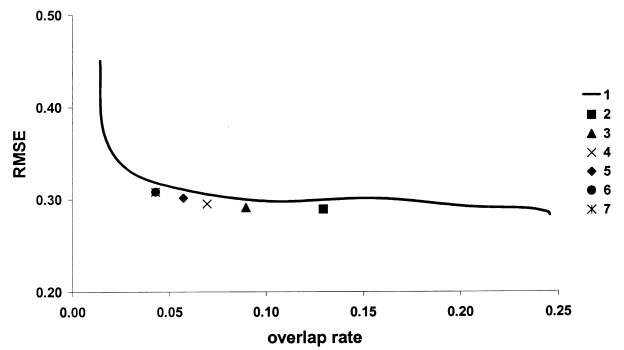
a) Uncorrelated parameters. Mean trait level equal to 0.



b) Uncorrelated parameters. Mean trait level equal to -.4.



c) Correlated parameters. Mean trait level equal to 0.



d) Correlated parameters. Mean trait level equal to -.4.

Figure 5. Relation, for the master item bank and the 6 different rotating banks, between overlap rate and RMSE for the two item banks and the two examinees' distributions.

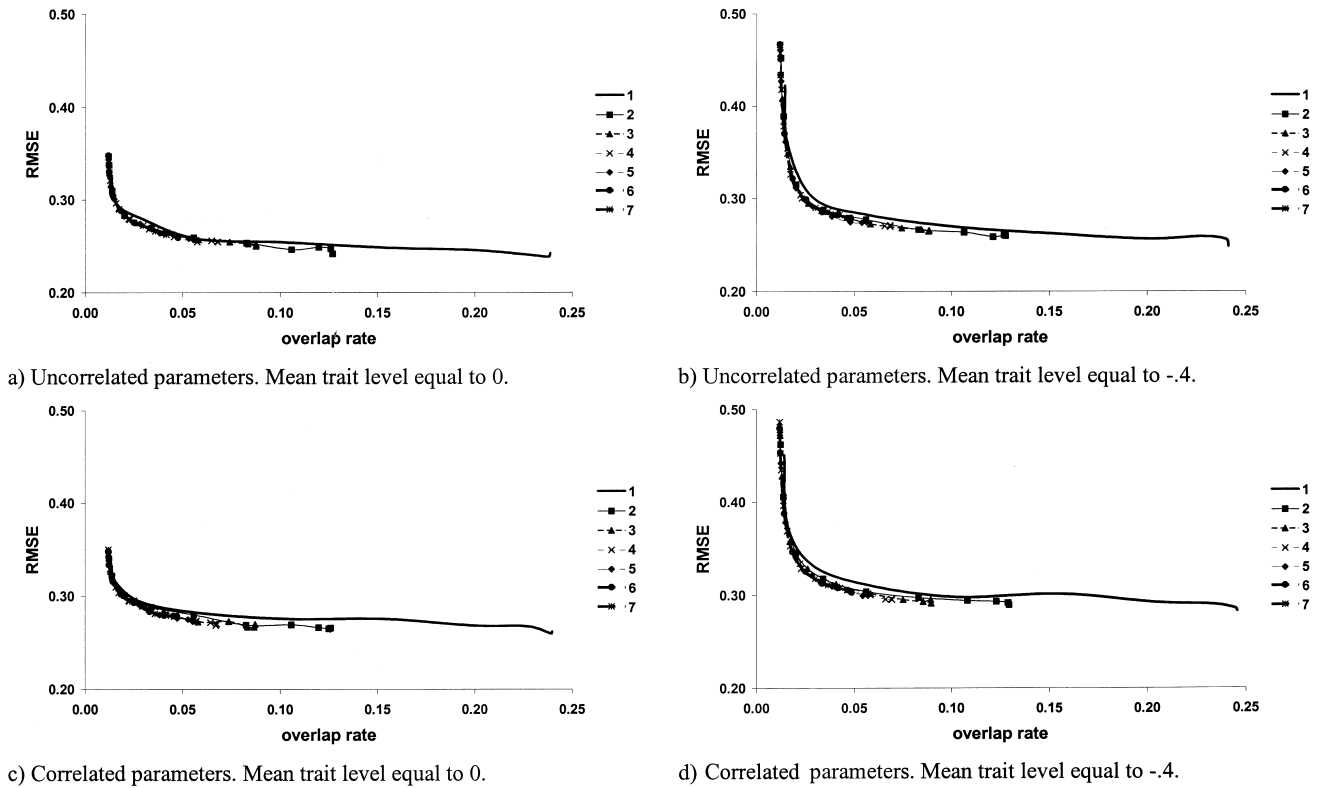


Figure 6. Relation, for the master item bank and the 6 different rotating banks combined with the SH method, between overlap rate and RMSE for the two item banks and the two examinees' distributions.

item banks slightly outperformed the results of the master item bank. The high degree of overlap between the different lines is an indicator of the small differences between conditions.

Discussion and Conclusions

Our aim was to compare three possible ways of increasing item bank security: the SH method, rotating item banks and the combination of both approaches. Given that increases in security involve losses in measurement accuracy, we set out to determine which of the three systems presents the least unbalanced trade-off between the two variables. To achieve this objective, we proposed and described a method that allows us to plot the relation between RMSE and overlap rate, allowing comparisons in one variable while maintaining the other constant. This method is useful in both research and applied contexts. In research, it allows more rigorous control of variables, avoiding the need to fix the r_{max} tentatively in order to allow comparison between methods. In applied contexts, it offers information that is simple to analyze, providing a guide to which levels of r_{max} control can be applied on an operational bank without surpassing a cost to measurement accuracy that is considered excessive.

Comparison between the functioning of a randomly generated bank divided into several sub-banks and that same

bank subjected to greater control on r_{max} shows that the three options are very similar, although the rotating option consistently outperformed the option of restricting r_{max} in the master bank. This is also the case when the rotating banks strategy is combined with the SH method. We have found that this results holds for two different kinds of item banks, with correlated and uncorrelated item parameters, and when the distribution of the examinees' trait level match the b parameter distribution and when those distributions do not match.

So far, we have only considered as relevant variables for the comparison between methods the accuracy of the trait level estimation and the risks for item bank security. Additional information for choosing between methods could be considered in applied contexts. For instance, given that the differences between alternatives are quite small, some CAT programs might prefer the SH method, as they may consider it easier to implement and be more familiar with a method proposed more than 20 years ago. From our perspective, the rotating strategy should be preferred, for reasons additional to those of the RMSE and overlap. First, item selection is faster when rotating, because fewer items are evaluated. Second, defining the incompatibilities between items is easier the smaller the size of the bank or sub-bank is. Imagine that your item bank has item enemies (items which, in the event one is administered, the other should not). To determine the matrix of item enemies is, clearly,

much harder when you have to consider a 2100*2100 cells, instead of 300*300 cells, for instance. Third, when we have been describing the security problems of the bank we have restricted ourselves to the problem of content spreading because of some examinees memorizing the items they receive and sharing them. Another possible risk source for the item bank is that someone, somehow, can get the whole file of item statements. When rotating, the CAT program loses just a part of the bank. If not rotating, the entire item bank is lost.

Given the results from the simulation study and the additional considerations just commented, our advise is to apply the rotating banks strategy whenever possible, combining it with a method for restricting r_{max} if needed.

References

- Ariel, A., Veldkamp, B. P., & van der Linden, W. J. (2004). Constructing rotating item pools for constrained adaptive testing. *Journal of Educational Measurement, 41*, 345-359.
- Barrada, J. R., Olea, J., & Ponsoda, V. (2007). Methods for restricting maximum exposure rate in computerized adaptive testing. *Methodology, 3*, 14-23.
- 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*.
- Barrada, J. R., Veldkamp, B. P., & Olea, J. (In press). Multiple maximum exposure rates in computerized adaptive testing. *Applied Psychological Measurement*.
- 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., & van der Linden, W. J. (2003). Optimal stratification of item pools in α -stratified computerized adaptive testing. *Applied Psychological Measurement, 27*, 262-274.
- Chang, H. H., & Ying, Z. (1999). α -Stratified multistage computerized adaptive testing. *Applied Psychological Measurement, 23*, 211-222.
- 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., & Lei, P. W. (2005). Controlling item exposure and test overlap in computerized adaptive testing. *Applied Psychological Measurement, 29*, 204-217.
- 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*.
- Davey, T., & Nering, N. (2002). Controlling item exposure and maintaining item security. In C. N. Mills, M. T. Potenza, J. J. Fremer, & W. C. Ward, (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 165-191). Mahwah, NJ: Erlbaum.
- 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.
- Lu, Y., & Hambleton, R. K. (2004). Statistics for detecting disclosed items in CAT environment. *Metodologia de las Ciencias del Comportamiento, 5*, 225-242.
- Mills, G. N., & Steffen, M. (2000). The GRE computer adaptive test: Operation issues. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 75-100). Boston: Kluwer Academic Press.
- 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). A sharing item response theory model for computerized adaptive testing. *Journal of Educational & Behavioral Statistics, 29*, 439-460.
- Stocking, M. L. (1994). *Three practical issues for modern adaptive testing item pools*. ETS Research Report No 94-05. Princeton, NJ: Educational Testing Service.
- Stocking, M. L., & Swanson, L. (1998). Optimal design of items banks for computerized adaptive tests. *Applied Psychological Measurement, 22*, 271-279.
- 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 & Behavioral Statistics, 28*, 249-265.
- van der Linden, W.J. (2005). *Linear models for optimal test design*. New York: Springer.
- 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.
- Veerkamp, W. & Glas, C.A.W. (2000). Detection of known items in adaptive testing with a statistical quality control method. *Journal of Behavioral and Educational Statistics, 25*, 373-389.
- Wainer, H. (2000). Rescuing computerized testing by breaking Zipf's law. *Journal of Educational & Behavioral Statistics, 25*, 203-224.
- Way, W. D. (1998). Protecting the integrity of computerized testing item pools. *Educational Measurement: Issues and Practice, 17*, 17-27.
- Wise, S. L., & Kingsbury, G. G. (2000). Practical issues in developing and maintaining a computerized adaptive testing program. *Psicologica, 21*, 135-155.

Received October 3, 2006

Revision received February 22, 2008

Accepted April 24, 2008