Item Selection Algorithms in Computerized Adaptive Test Comparison Using Items Modeled with Nonparametric Isotonic Model

Mario Luzardo

Abstract A computerized adaptive test (CAT) is used in this paper where the item bank is calibrated by using the nonparametric isotonic model proposed by Luzardo and Rodríguez (Quantitative psychology research. Springer International Publishing, Switzerland, pp. 99-108, 2015). The model is based on the estimation of the inverse of the item characteristic curves (ICC), and it uses a two-stage process. First, it uses the Ramsay nonparametric estimator of the ICC (Ramsay In Psychometrika 56:611–630, 1991) and then it estimates the density function of the inverse ICC by using Ramsay's estimator. By integrating the density function and then symmetrizing it, we obtain the result. Xu and Douglas (Psychometrika 71:121–137, 2006) studied the possibility of using Ramsay's nonparametric model in a CAT. They explored the possible methods of item selection but they did not use Fisher's maximum information method because the derivatives of the ICC may not be estimated well. We present, for the isotonic model, a suitable way to estimate the derivatives of the ICCs and obtain a formula for item information that allows us to use the maximum information criterion. This work focuses on comparing three methods for selecting items in the CAT: random selection, the maximum Fisher information criterion with the isotonic model, and the Kullback-Leibler information criterion.

Keywords Isotone IRT nonparametric model · Kullback-Leibler information · Computerized adaptive test

1 Introduction

Nonparametric item response models have been an alternative to parametric item response models, especially when it comes to finding a flexible model for ICC modelling. However, a common problem is how to make CAT administration and, in particular, automatic item selection operational.

M. Luzardo (✉)
School of Psychology, University of the Republic, Montevideo, Uruguay
e-mail: mluzardo@psico.edu.uy

Eastern Regional University Center, University of Republic, Maldonado, Uruguay

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings in Mathematics & Statistics 265, https://doi.org/10.1007/978-3-030-01310-3_9

Xu and Douglas (2006) explored the possibility of applying CAT by using Ramsay's nonparametric model. Under this model, the usual methods for estimating the ICC derivative do not work properly and the derivative may be negative for some values of the ability. This means it is impossible to use the maximum Fisher information criterion when choosing the items to be managed. Xu and Douglas (2006) propose as alternative the use of procedures based on Shannon entropy (Cover & Thomas, 1991; Shannon, 1948) and Kullback-Leibler information (Chang & Ying, 1996), since the implementation of these procedures does not require ICC derivatives. In addition, when the test size is large enough, they are equivalent to the maximum Fisher information criterion. The authors used a simulation study to show that both procedures work properly and have very similar outcomes.

In this paper we will show that when the nonparametric isotonic model is used to estimate the ICCs, their derivatives can be calculated in a simple way, and they can be used to estimate the Fisher information for each item. Our aim is therefore to compare this new approach with those proposed by Xu and Douglas (2006). Since the Kullback-Leibler procedure and that based on the Shannon entropy produce very similar results, we will only use Kullback-Leibler. Our main intention is to compare performances in the case of small test sizes, since the Kullback-Leibler and the Shannon entropy procedures are asymptotically equivalent to the maximum information criterion.

2 One-Dimensional Isotonic Model

The isotonic model presented in Luzardo & Rodríguez (2015) estimates the ICC in two stages. The first stage uses the Ramsay model (1991) as a preliminary estimate of the ICC, and the second obtains the isotonic estimator.

Let X be a dichotomous item and assume that $P(\theta)$ is the probability that a subject with ability θ will respond to item X correctly. As the random variable X is Bernoulli, it follows that $P(\theta) = E(X|\Theta = \theta)$, that is, the ICCs match a conditional expectation. On this basis, Ramsay estimated the ICCs by means of a nonparametric kernel regression estimator.

Let us assume that N subjects with a latent trait $\theta_1 \ldots \theta_N$ respond to n dichotomous items. Let us denote X_{ij} as the binary response of subject i to item j $(i = 1,\ldots, N\ j = 1,..,n)$. The kernel smoothing estimator of $P_j(\theta)$ is

$$\widehat{P_j(\theta)} = \frac{\sum_{i=1}^{N} X_{ij} K_h\left(\hat{\theta}_i - \theta\right)}{\sum_{i=1}^{N} K_h\left(\hat{\theta}_i - \theta\right)} \tag{1}$$

where the bandwidth h contemplates the trade-off between the variance of the estimator and the bias. Function K is a kernel and $K_h(\theta_i - \theta) = \frac{1}{h} K\left(\frac{(\theta_i - \theta)}{h}\right)$. In Eq. (1), $\hat{\theta}_i$ is the estimator of the i-th subject's ability. These estimates can be easily calculated by converting the empirical distribution of the sum of the subjects' scores to the scale determined by the distribution of the ability.

We will take—with no loss of generality—θ to have a uniform distribution in [0,1]. This assumption is justified by the non-identifiability of the scale. Let us assume that the distribution of the actual trait τ is $F(\tau)$ and let us consider a specific item with a strictly increasing ICC, which we will denote as $P(\tau)$.

If we change the variable $\theta = F(\tau)$, the function $P^*(\theta) = P\left(F^{-1}(\theta)\right) = P(\tau)$ is also the ICC of that item. It is clear that the distribution of θ is uniform in [0,1] and $P^*(\theta)$ is increasing.

Note that if $U_1, \ldots, U_T$ is a sample of independent random variables with a uniform distribution on the interval [0,1], then $\frac{1}{Th_d}\sum_{t=1}^{T} K_d\left(\frac{P^*(U_t)-u}{h_d}\right)$ is an estimator of the density of the random variable $P^*(U)$, where K_d is a kernel and h_d a bandwidth.

The density of $P^*(U)$ is $P^{*-1'}(u)\mathbb{I}_{[P^*(0),P^*(1)]}(u)$, where $\mathbb{I}$ is the indicator function. Then, $\frac{1}{Th_d}\int_{-\infty}^{\theta}\sum_{t=1}^{T} K_d\left(\frac{P^*(U_t)-u}{h_d}\right)du$ is a consistent estimator of P^{*-1} in θ (Dette, Neumeyer, & Pilz, 2006).

In order to apply the above property to our problem, let us consider a kernel K_r a bandwidth h_r, and a grid $\frac{1}{T}, \ldots, \frac{t}{T}, \ldots, 1$. Then, the Ramsay estimator of the ICC in each score is

$$\widehat{P^R}\left(\frac{t}{T}\right) = \frac{\sum_{i=1}^{N} K_r\left(\frac{\frac{t}{T}-\hat{\theta}_i}{h_r}\right)X_i}{\sum_{i=1}^{N} K_r\left(\frac{\frac{t}{T}-\hat{\theta}_i}{h_r}\right)} \tag{2}$$

Based on the above, the isotonic estimator of the inverse of the ICC in θ is:

$$\widehat{P^{*-1}(\theta)} = \frac{1}{Th_d}\int_{-\infty}^{\theta}\sum_{t=1}^{T} K_d\left(\frac{\widehat{P^R}\left(\frac{t}{T}\right)-u}{h_d}\right)du \tag{3}$$

The estimator $\widehat{P^*}$ is obtained by the reflection of $\widehat{P^{*-1}}$ with respect to the bisector of the first quadrant.

3 Item Selection Method Through Maximum Information

The maximum information procedure is based on the fact that when the maximum likelihood method is used to estimate the ability, the test information is inversely proportional to the estimation error of θ. It is therefore reasonable to present in the next step the item that will maximize the accumulated information. This procedure will be adopted in this article. It is therefore necessary to be able to correctly estimate the derivative of the ICC of the nonparametric isotonic model.

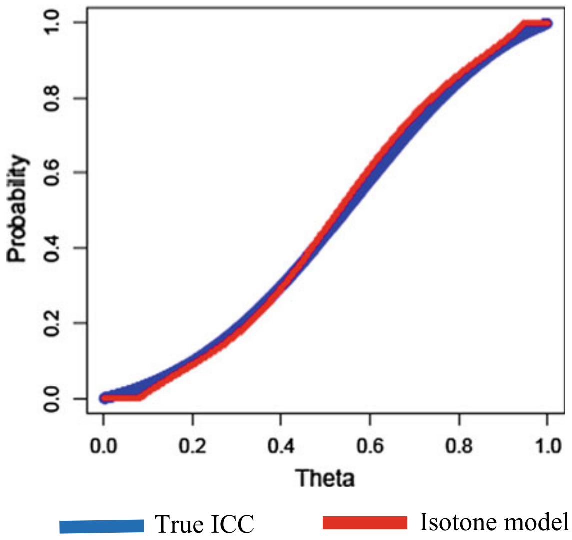


Fig. 1 True ICC and estimated ICC using the isotone model

In our case, we easily obtain a simple expression for the derivative of the ICC which is smooth and always positive by applying the inverse function derivative theorem.

$$P^{*'}(\theta) = \frac{1}{\left(P^{*-1}\right)'(P^{*}(\theta))} = \frac{Th_d}{\sum_{t=1}^{T} K_d\left(\frac{P^{*}(U_t)-P^{*}(\theta)}{h_d}\right)} \tag{4}$$

Figure 1 shows the true ICC and the estimated ICC by means of the isotonic model, and Fig. 2 shows the derivatives of this ICC.

Now, on the basis of (4), we can estimate the information function of item j through:

$$\widehat{I_j(\theta)} = \frac{\left(\frac{\partial \widehat{P_j^{*}}(\theta)}{\partial\theta}\right)^2}{\widehat{P_j^{*}}(\theta)\left(1-\widehat{P_j^{*}}(\theta)\right)} = \frac{\left[\frac{Th_d}{\sum_{t=1}^{T} K_d\left(\frac{\widehat{P^{*}\left(\frac{t}{T}\right)}-P^{*}(\theta)}{h_d}\right)}\right]^2}{\widehat{P_j^{*}}(\theta)\left(1-\widehat{P_j^{*}}(\theta)\right)} \tag{5}$$

Information estimation works very well on values where item information is maximum, having distortions when we move away from that value. Our interest is focused on a setting where information is maximum, so outside that neighborhood, we can estimate information using a linear model. Figure 3 shows the information function estimate.

Fig. 2 True derivative and estimated derivative of the ICC

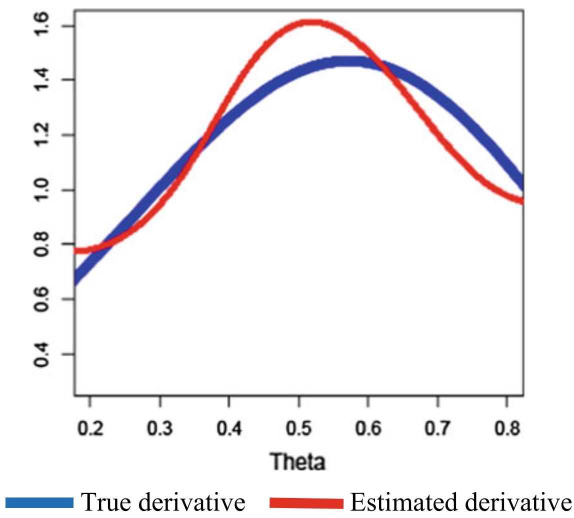
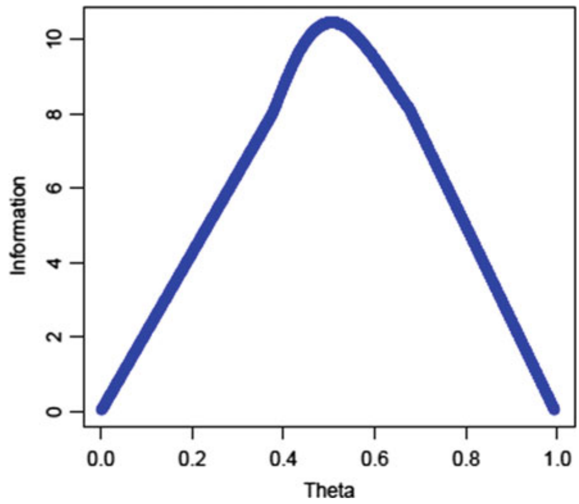


Fig. 3 Estimated information function



The maximum likelihood method will be used to estimate the ability. If $P^*(\theta)$ is the ICC when the ability follows a uniform distribution, θ will be estimated in step k through

$$\hat{\theta}_i = argmax \prod_{j=1}^{k} P^*(\theta)^{X_{ij}} \left(1 - P^*(\theta)\right)^{1-X_{ij}} \tag{6}$$

4 Item Selection Method Using Kullback-Leibler

This divergence proposed by Kullback and Leibler (KL) (1951) measures the discrepancy between two measures of probability. On the basis of this, Chang and Ying (1996) define a measure of global information for use in CAT.

If P and Q are two probability measures over Ω, and if $\frac{dQ}{dP}$ is the Radon-Nikodym derivative of Q with respect to P, the Kullback Leibler divergence is defined as:

$$KL(P\|Q) = -\int_{\Omega} ln\frac{dQ}{dP}dP \tag{7}$$

In particular, if μ is a measure over Ω, such that f and g are densities of P and Q with respect to μ, then

$$KL(P\|Q) = \int_{\Omega} \frac{f\ \ln\ f}{g}d\mu \tag{8}$$

If we consider the maximum likelihood estimate in a parametric family $f(\theta, x)$, and $f(\theta_0, x)$ as the true density, then:

$$KL\left(f_{\theta_0}\|f_{\theta}\right) = \int f(\theta_0, x)\, ln\frac{f(\theta_0, x)}{f(\theta, x)}dx \tag{9}$$

Chang and Ying (1996) define the Kullback Leibler information for item j and subject i as

$$KL_j(\theta\|\theta_i) = E\left[ln\frac{L_j(\theta_i|X_{ij})}{L_j(\theta|X_{ij})}\right] = P_j(\theta_i)ln\frac{P_j(\theta_i)}{P_j(\theta)} + \left(1 - P_j(\theta_i)\right)ln\frac{1 - P_j(\theta_i)}{1 - P_j(\theta)} \tag{10}$$

In the context of CATs, if $\hat{\theta}_k$ is the maximum likelihood estimator of θ, after k items have been responded to, then the global information index $GKL_j\left(\hat{\theta}_k\right)$ is obtained by taking the average of the discrepancy $KL_j\left(\theta\|\hat{\theta}_k\right)$ in the interval centered on $\hat{\theta}_k$, that is, if $\epsilon_k > 0$,

$$GKL_j\left(\hat{\theta}_k\right) = \int_{\hat{\theta}_k-\epsilon_k}^{\hat{\theta}_k+\epsilon_k} K_j\left(\theta\|\hat{\theta}_k\right)d\theta \tag{11}$$

The sequence $\epsilon_k \to 0$ with k. Chang and Ying (1996) recommend $\epsilon_k \propto k^{-\frac{1}{2}}$ so that the interval $\left(\hat{\theta}_k - \epsilon_k, \hat{\theta}_k + \epsilon_k\right)$ will contain the actual value of the ability. Based on the above, the item to be chosen for step (k + 1) will be the one with the greatest GKL, which has not been applied yet.

5 Simulation Study

The objective of this study was to compare three ways of selecting items in the CAT. The selection methods implemented are: the Kullback-Leibler procedure, the information-based procedure using the isotonic estimation, and random selection of items. The ability was estimated by using maximum likelihood and considering the nonparametrically estimated ICCs. Additionally, the ability for random item selection was estimated by maximum likelihood, when the ICC is estimated parametrically.

A bank of 700 items was built whose ICCs followed the two-parameter logistic model (2PL). The discrimination parameters of the items were simulated from a uniform distribution [0.75, 2.5] and the difficulty parameters were simulated from a uniform distribution [−2, 2].

To estimate the ICCs, the responses of 5000 subjects were simulated. The abilities were assumed to follow a standard normal distribution. On the basis of the responses, we used the isotonic estimator with Gaussian kernels K_r and K_d.

The bandwidths used were $h_r = (5000)^{\left(-\frac{1}{5}\right)} = 0.18$, and a robust estimate for $h_d = 0.9(5000)^{\left(-\frac{1}{5}\right)}\min\left(sd, \frac{Q_3-Q_1}{1.364}\right)$, where the deviation and the quartiles refer to the Ramsay's estimator of the ICC for each item.

For the CAT, 5000 subjects were generated whose traits had a uniform distribution on the interval [0,1]. A test of 50 items in length was applied for each of the methods and the procedures for each subtest of 5, 10, 20, 30, 40 and 50 items in length were assessed. The different procedures were compared to root mean squared error (RMSE) and bias across the simulations. The RSME and bias were computed for each subtest through:

$$RMSE = \left(\frac{\sum_{i=1}^{N}\left(\hat{\theta}_\iota - \theta_i\right)^2}{N}\right)^{\left(\frac{1}{2}\right)} \tag{12}$$

$$BIAS = \frac{\sum_{i=1}^{N}\left(\hat{\theta}_\iota - \theta_i\right)^2}{N} \tag{13}$$

Also, the RSME and bias of the estimators were calculated with a certain value of θ through:

$$RMSE(\theta) = \left(\frac{\sum_{i\epsilon I(\theta)}\left(\hat{\theta}_\iota - \theta_i\right)^2}{\#I(\theta)}\right)^{\left(\frac{1}{2}\right)} \tag{14}$$

$$BIAS(\theta) = \frac{\sum_{i\epsilon I(\theta)}\left(\hat{\theta}_\iota - \theta_i\right)^2}{\#I(\theta)} \tag{15}$$

where $I(\theta) = \{i : \theta_i = \theta,\ 1 \le i \le N\}$

6 Results

Table 1 presents the average root mean square error for the different methods and for different subtest lengths. This table shows how the procedures based on the information estimated from the isotonic ICC and the Kullback-Leibler method work in a similar way. In addition, these methods are better than the random selection of items, and the estimation of the ability is based on the isotonic nonparametric model when fewer than 30 items are administered. They are also always better than random selection and ability estimation using 2PL model. Table 2 shows a similar behavior for the bias.

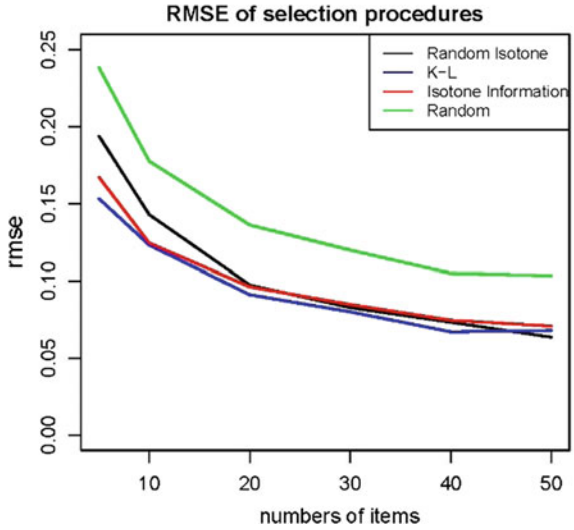
Figure 4 graphically shows how the RMSE stabilizes after a test length of 20 items for the nonparametric isotone model and Kullback-Leibler procedures. The Fig. 5 shows that the bias is also stabilized.

Table 1 Average root mean square error

Selection rule	Number of items					
	5	10	20	30	40	50
Random isotone	0.203	0.147	0.108	0.085	0.074	0.074
K-L	0.154	0.125	0.098	0.085	0.076	0.074
Isotone Information	0.154	0.123	0.098	0.084	0.080	0.074
Random 2PL	0.242	0.181	0.142	0.116	0.105	0.103

Calculated on the 5000 subjects

Fig. 4 RMSE of selection procedures



mluzardo@psico.edu.uy

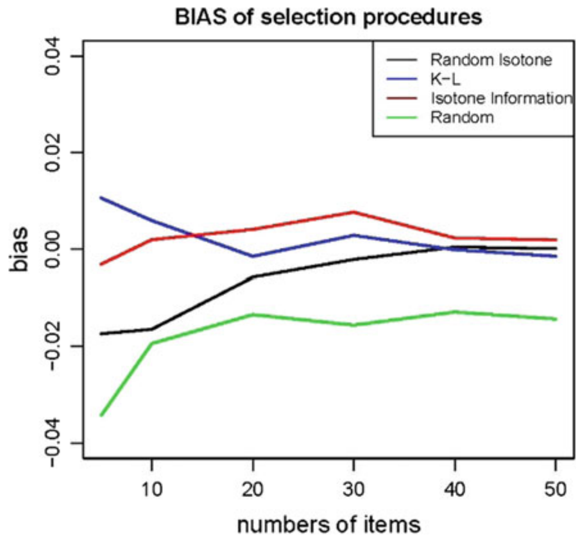


Fig. 5 Bias of selection procedures

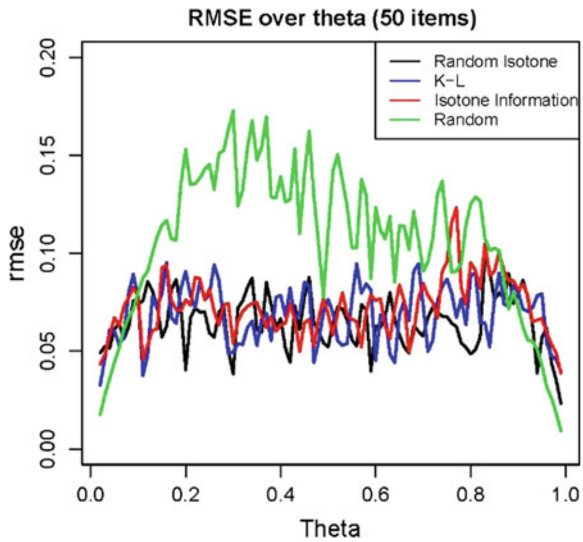


Fig. 6 RMSE over theta

An analysis of the RMSE(θ) finds that the same behavior is obtained for all θ, with an equivalence of the Kullback-Leibler method and the nonparametric isotonic method and the latter's superiority over random selection using de 2PL model procedure (Fig. 6).

When the bias is analyzed globally and as a function of θ, both methods behave appropriately (Fig. 7).

Fig. 7 BIAS over theta

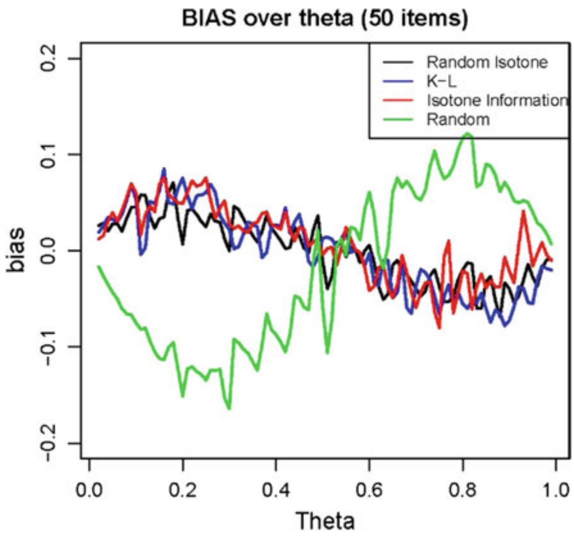


Table 2 Bias

Selection rule	Number of items					
	5	10	20	30	40	50
Random isotone	0.012	−0.013	0.003	−0.002	0.001	0.001
K-L	−0.003	−0.009	0.002	0.004	0.0001	0.001
Isotone Information	0.0004	0.003	0.007	0.005	0.005	0.005
Random 2PL	0.002	−0.014	−0.011	−0.012	−0.01	−0.01

Calculated on the 5000 subjects

7 Discussion

The procedure based on estimating information through the isotonic model quickly converges to the actual trait, stabilizing after 20 items. The performance of the procedure presented based on the isotonic model is similar to that of KL in terms of root mean square error, and simpler to implement. It is also observed that both adaptive procedures work better than random selection of items in terms of root mean square error.

It would be wise to expand some studies that would extend these results, for example by studying the rate of exposure of the items, which it has been omitted in this work.

References

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

Cover, H. H., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Dette, H., Neumeyer, N., & Pilz, K. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli, 12*(3), 469–490.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694.

Luzardo, M., & Rodríguez, P. (2015). A nonparametric estimator of a monotone item characteristic curve. In *Quantitative Psychology Research* (pp. 99–108). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-19977-1_8.

Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika, 56,* 611–630.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27,* 379–423.

Xu, X., & Douglas, J. (2006). Computerized adaptive testing under nonparametric IRT models. *Psychometrika, 71,* 121–137.