Journal of Educational and Behavioral Statistics
December 2010, Vol. 35, No. 6, pp. 693–713
DOI: 10.3102/1076998610375834

Bayesian Estimation of the Logistic Positive Exponent IRT Model

Heleno Bolfarine
Universidade de Sao Paulo

Jorge Luis Bazan
Pontificia Universidad Católica, Perú

A Bayesian inference approach using Markov Chain Monte Carlo (MCMC) is developed for the logistic positive exponent (LPE) model proposed by Samejima and for a new skewed Logistic Item Response Theory (IRT) model, named Reflection LPE model. Both models lead to asymmetric item characteristic curves (ICC) and can be appropriate because a symmetric ICC treats both correct and incorrect answers symmetrically, which results in a logical contradiction in ordering examinees on the ability scale. A data set corresponding to a mathematical test applied in Peruvian public schools is analyzed, where comparisons with other parametric IRT models also are conducted. Several model comparison criteria are discussed and implemented. The main conclusion is that the LPE and RLPE IRT models are easy to implement and seem to provide the best fit to the data set considered.

Keywords: *achievement; assessment; item response theory (IRT); mathematics education*

1. Introduction

In successive papers, Samejima (1995, 1997a, 1997b, 2000) has presented the derivation of an item response theory (IRT) model, namely, the logistic positive exponent (LPE) family of models, which consider an asymmetric item characteristic curve (ICC) that can be appropriate in many situations, because, as shown in those papers, symmetric ICCs as the normal ogive and the Logistic IRT can lead to an intrinsic contradiction in the philosophy of ordering individuals on the ability scale. A characteristic of a symmetric ICC is that it treats both correct and incorrect answers symmetrically, which results in a logical contradiction in ordering examinees on the ability scale.

In addition, LPE models include the item task complexity because the third item parameter is different from item discrimination and item difficulty parameters, which determine the single principle of ordering individuals on the ability scale.

The LPE model has a high degree of substantive validity and inner consistency in ordering individuals (Samejima, 1997b) and the point-symmetric (logistic) model is treated as one of the infinitely many models in the family, but LPE seems to provide more appropriate ICCs. Thus, the LPE model includes as special case the logistic IRT model so that it is a more flexible model allowing symmetric and asymmetric ICCs for the items in a test. In addition, as proved in Samejima (2000), the contradiction in the rank order of response patterns does not exist in LPE models.

In this article, we introduce Bayesian estimation to the LPE model. In addition, another IRT model is introduced. As will be shown later, this new model is a reflection of the LPE model and is named here as RLPE. Both models can be considered as skewed logistic IRT models and have as its main characteristic a point-asymmetric ICC. As also seen later, both models are derived by considering two links proposed by Prentice (1976).

Bayesian estimation will be developed using the MCMC methodology and the WinBUGS software, which can be used for simulating from the posterior distributions of item parameters and latent variables.

The main objective of this article is to offer a clear presentation of Bayesian estimation via MCMC for the two skewed logistic IRT models considered. The article is organized as follows. In Section 2, we introduce the LPE IRT model by considering a particular ICC based in a skew-logit link. Moreover, a new skewed logistic IRT model is proposed by considering another ICC based in other skew-logit link. In Section 3, we discussed the inference for the models considered and we deal with Bayesian inference including several model comparison criteria. In Section 4, we illustrate the methodology with simulated data sets (Samejima, 2000). An example is given in Section 5, illustrating the usefulness of the approach in comparing it with other parametric IRT models using a real data set. To choose the model that fits the data better, we consider the deviance information criterion (DIC) as presented in Spiegelhalter, Best, Carlin, and van der Linde (2002) as well as other model comparison criteria. Finally, we discuss possible extensions of the model proposed.

2. Skewed Logistic IRT Models

2.1. Symmetric ICCs

IRT models to dichotomous item responses assume that the sequence of binary random variables $\{Y_{ij} : 1 \le i \le n;\ 1 \le j \le k\}$ associated with item responses are conditionally independent, given θ_i, the latent variable associated with the ability or latent trait for individual i. It is considered that $Y_{ij} = 1$, if subject i correctly answers item j, and $Y_{ij} = 0$ otherwise. The response pattern of person i is written as $Y_i = (Y_{i1}, \ldots, Y_{ik})$. It also is assumed that the probability of the event $Y_{ij} = 1$ (correct response), namely, p_{ij}, can be written as

$$p_{ij} = P\left[Y_{ij} = 1 \middle| \theta_i, a_j\, b_j\right] = F\left(m_{ij}\right), \tag{1}$$

where F is called the ICC, and

$$m_{ij} = a_j\left(\theta_i - b_j\right), \quad i = 1, \ldots, n, \quad j = 1, \ldots, k \tag{2}$$

is a latent linear predictor where a_j and b_j are parameters associated with the items (denominated discrimination and difficulty parameters, respectively).

Two known cases of ICCs follow by considering in (Equation 1) the cumulative distribution function (cdf) of the standard normal distribution and the cdf of the standard logistic distribution. Such models usually are called the normal ogive IRT model and Logistic IRT model, respectively, denoted here as 2P and 2L IRT models. When $a_j = 1$ in (Equation 2), we obtain models 1P and 1L by considering only item difficulty parameters. In addition, models 3P and 3L are obtained when we consider $p_{ij} = c_j + \left(1 - c_j\right)F\left(m_{ij}\right)$ in Equation 1, where c_j is a pseudo-chance-level parameter or guessing parameter, indicating that the probability of correct response is greater than zero even for those with very low trait levels.

In the context of generalized linear models, the inverse function of $F(.)$ in (Equation 1) is called the link function. The models can also be named as probit and logit IRT models, respectively, by emphasizing the link considered. A special feature of both models is the symmetric nature of the probit and logit link or of the corresponding ICCs used. These models also are named point-symmetric models (Samejima, 1997a).

However, as emphasized in Chen, Dey, and Shao (1999), in the context of binary regression, symmetric links do not always provide good fit for some data sets. This is especially true, when the probability of a given binary response approaches zero at a different rate then it approaches one. As pointed out by several authors, misspecification of the link function can yield substantially biased mean response estimates (Czado & Santner, 1992).

2.2. Some Asymmetric ICCs in IRT

A variety of asymmetric ICCs or asymmetric links have been proposed for the binary regression models (see Bazan, Bolfarine, & Branco, 2010), but hardly the two have been used in IRT models (the skew-probit model proposed by Bazan, Branco, & Bolfarine, 2006, and the model proposed by Samejima, 1997a), including the 1P, 2P and 1L, 2L models as particular cases, respectively.

However, different asymmetric ICCs to IRT models can be considered by taking in (Equation 1) the cdf of an asymmetric distribution. A very popular example of this situation is the well-known and widely used complementary log–log link, where the cdf of the Gumbel distribution is considered. However, the cdf of other distributions such as the Weibull and log-normal distributions can also be used to define new IRT models. In such cases, the cdf is completely specified, and it does not depend on any unknown parameter and no relationship between them and the usual symmetric links are established, which can be a nuisance restriction. Other less restrictive ICCs to IRT models can be obtained when considering the following cdfs:

$$F_1(x) = 1 - (1 + e^x)^{-\lambda} \quad \text{and} \quad F_2(x) = (1 + e^{-x})^{-\lambda}, \quad \lambda > 0. \tag{3}$$

In spite of the fact that there is no well agreed name for the first cdf, Achen (2002) has named it as the standard scobit distribution and the second cdf corresponds to the standard burr Type II distribution (Johnson, Kotz, & Balakrishnan, 1994). Corresponding links using distributions $F_1(.)$ and $F_2(.)$ were proposed in Prentice (1976) and were popularized in the statistical literature by Aranda-Ordaz (1981) and in the econometric literature by Nagler (1994) and Achen (2002). These links are skewed modifications of the logit link and are here termed *scobit* and *power logit*, respectively, and include the logit link as special case by considering the parameter $\lambda = 1$. In general, if we define $Y = \mu + \sigma X$, we say that a variable $Y \sim$ scobit (μ, σ) ("$\sim$" meaning "distributed as") or $Y \sim$ burr_{II} (μ, σ). The corresponding probability density function (pdf) are given by

$$f_1(y) = \frac{\lambda\left[\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{\sigma\left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda+1}} \quad \text{and} \quad f_2(y) = \frac{\lambda\left[\exp\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}}{\sigma\left[1 + \exp\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda+1}},$$

respectively, where μ is a location parameter and σ is a scale parameter. For example, if $Y \sim \text{burr} - \text{II}\ (\mu, \sigma)$ then $\frac{Y-\mu}{\sigma} \sim \text{burr} - \text{II}(0, 1)$.

Note that $F_1(-y) \neq 1 - F_1(y)$ or $F_2(-y) \neq 1 - F_2(y)$ and then F_1 and F_2 are not point-symmetric but $F_1(-y) = 1 - F_2(y)$, and thus, the burr-II and the scobit distributions are distinct, though closely related because one is the reflection of the other.

2.3. LPE and the Reflection of LPE IRT Models

The LPE IRT model proposed in Samejima (1997a, 1997b, 2000) can be obtained by considering the power logit link or, equivalently, when the Burr type II distribution $F_2(.)$ is considered as an ICC in Equation 1. Moreover, the formulation of the LPE model essentially implies that the Logistic IRT model is nested within the LPE model. However, another interesting IRT model can be obtained when the scobit link or the scobit distribution $F_1(.)$ is considered as an alternative ICC. This other IRT model is denominated here as the reflection of the LPE IRT model, namely, the RLPE model. In general, we say that LPE and RLPE IRT models are s*kewed logistic* IRT models and are obtained by replacing Equation 1 by

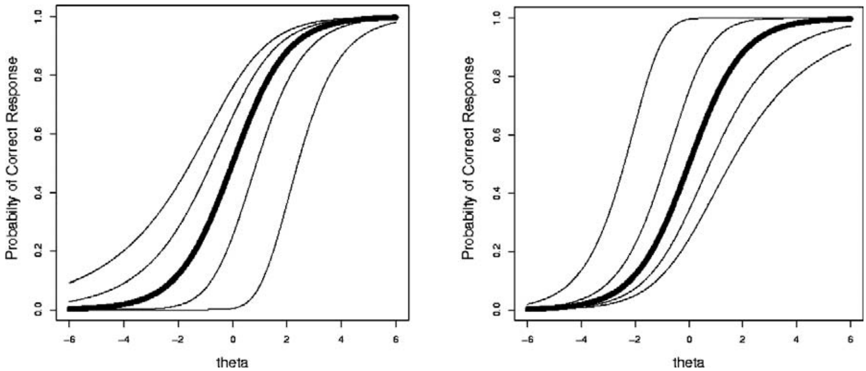

FIGURE 1. *Probability curves for $\lambda = 0.4, 0.6, 1, 2, 8$ in LPE (left) and for $\lambda = 8, 6, 1, 0.6, 0.4$ in Reflection Logistic Positive Exponent (RLPE) (right) models considering different ranges for θ and $a = 1$, $b = 0$. For $\lambda = 1$, the logistic model is obtained.*

$$p_{ij} = P\left[Y_{ij} = 1 \middle| \theta_i, a_j\, b_j\right] = F_{\lambda_j}\left(m_{ij}\right), \tag{4}$$

where F_{λ_j} is the cdf $F_2(.)$ (or $F_1(.)$) indexed with λ_j and evaluated at m_{ij} given by Equation 2. When $F_2(.)$ is considered, we have the LPE model and when $F_1(.)$ is considered, we have the RLPE model.

The LPE IRT model can also be obtained within Samejima's framework by considering $p_{ij} = L\left(m_{ij}\right)^{\lambda_j}$, with $L(.)$ as the cdf of the standard logistic function and $\lambda_j > 0$ is the shape parameter associated with the *j*th item, providing asymmetric ICCs and including the Logit IRT model as a special case when $\lambda_j = 1$. Hence, the LPE model is as a generalization of the logit link, which follows by introducing a shape parameter associated with the item, that is, it is interpreted as a *penalization* item parameter and can play an important role in testing, as emphasized in Samejima (1997, 2000).

Figure 1 depicts different probability curves or ICCs for the LPE and RLPE models using different values for θ, for an item with $a = 1$ and $b = 0$. For $\lambda = 1$, the ICC corresponds to the logistic IRT model and for $\lambda < 1$ (or $\lambda > 1$). The ICC corresponding to LPE model is generally above (below) the ICC corresponding to the logistic IRT model within a range of ability values. Note also that for each value of λ, RLPE ICC is a reflection of the LPE ICC, and thus for $\lambda < 1$ (or $\lambda > 1$), the corresponding ICC is generally below (above) the corresponding ICC for the logistic IRT.

For the LPE model, the parameter λ typically is called the acceleration parameter (Samejima, 1995), in the sense that this will accelerate the point (value) theta at which the slope or discrimination power of the ICC becomes greatest. For higher values of λ, we have a change point in the ICC more to the right.

In contrast, for the RLPE model, λ can be called the deceleration parameter because it has now an opposing behavior. That is, for higher values of λ, we have a change point more to the left.

In LPE, the ICC follows the principle of penalizing failure in solving an easier item when $0 < \lambda < 1$ (also it is considered that the item is less complex), whereas it follows the opposing principle; that is, greater credit is given for solving a more difficult item when $\lambda > 1$ (also, it is considered that the item is more complex).

It is expected that the ICC for a less complex task assumes higher probabilities of success than that for a more complex task, "By task complexity we mean the level of demands in a task" (Samejima, 1997a, p. 483).

A task is considered more complex if it requires an individual to perform or successively pass each of many sequential subprocesses to solve successfully the complete problem. It is expected that, if a process is tougher and/or contains a larger number of sequential subprocesses, the conditional probability, given the latent trait or ability, for performing the process successfully will become small. As is indicated in Samejima (2000), such differences in the conditional probability are expected to become more pronounced for lower ability levels. For convenience, let us call this aspect of an item complexity, as distinct from item difficulty, and taking this item complexity into account, the eventual ICC is asymmetric and depends on how many and how tough sequential subprocesses are involved in solving the problem. Note that the word sequential is used by Samejima in a very broad sense, and subprocesses may be either serial or parallel as is detailed in Samejima (1995).

Because that RLPE is a reflection of LPE, the principle of greater credit in solving an easier item is expected when the item is complex, in this case when $0 < \lambda < 1$, and penalize failure is given for solving more difficulty item or when item is less complex ($\lambda > 1$).

In LPE, if $\lambda < 1$, then even individuals with very low ability levels have substantially high probabilities to pass the item. In this case, if a test consists of items with common values of a, $\lambda < 1$ and different values of b, individuals failing to solve easier items are penalized. However, when $\lambda > 1$, then even individuals with high ability levels have a substantially low probability to pass the item. In this case, if a test consists of items with common values of a, $\lambda > 1$ and different values of b, individuals succeeding in solving more difficult items are rewarded.

In RLPE, if $\lambda < 1$, then individuals presenting very low ability levels have substantially low probabilities of failing the item. In this case, if a test consists of items with common a values, $\lambda < 1$ and different b values, then individuals with low abilities correctly solving easier items are rewarded. Alternatively, if $\lambda > 1$, then individuals with high ability levels have a substantially high probability of failing the item. In this case, if a test consists of items with common values of a, $\lambda > 1$ and different b values, then credits are given to fail in solving more difficult items.

3. Inference

3.1. Likelihood Function Versions

The likelihood function for the skewed logistic IRT class (or family) of models indexed by λ_j is given by

$$L(\beta, \theta|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n}\prod_{j=1}^{k}\left[F_{\lambda_j}\left(m_{ij}\right)\right]^{y_{ij}}\left[1 - F_{\lambda_j}\left(m_{ij}\right)\right]^{1-y_{ij}}, \tag{5}$$

where $\beta = (\mathbf{a}', \mathbf{b}')'$, $\mathbf{a} = (a_1, \ldots a_n)'$, $\mathbf{b} = (b_1, \ldots b_n)'$, m_{ij} is the latent linear predictor in Equation 2 and $F_{\lambda_j}(m_{ij})$ is the cdf $F_2(.)$ (or $F_1(.)$) in Equation 4, indexed with λ_j and evaluated at m_{ij}.

In this section, we present a complete data likelihood function for the skewed logistic IRT models, so that we start with an important alternative representation.

Proposition. The skewed logistic models, as defined before, can be equivalently written as

$$y_{ij} = I\left(s_{ij} > 0\right) = \begin{cases} 1, & s_{ij} > 0 \\ 0, & s_{ij} \leq 0 \end{cases}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k, \tag{6}$$

where $s_{ij} \sim scobit\left(m_{ij}, 1\right)$ if the LPE model is defined, or $s_{ij} \sim burr - II\left(m_{ij}, 1\right)$ if the RLPE model is defined, with $I(.)$ as the usual indicator function.

Proof. Note that for the LPE model

$$P\left(Y_{ij} = 1\right) = P\left(s_{ij} > 0\right) = 1 - P\left(s_{ij} \leq 0\right) = 1 - F_1\left(-m_{ij}\right)$$

$$= 1 - \left[1 - \left(1 + \exp\left(-m_{ij}\right)^{-\lambda j}\right)\right] = \left[1 + \exp\left(-m_{ij}\right)^{-\lambda j}\right] = F_2\left(m_{ij}\right)$$

and

$$P\left(Y_{ij} = 0\right) = P\left(s_{ij} < 0\right) = F_1\left(-m_{ij}\right) = 1 - \left[1 + \exp\left(-m_{ij}\right)^{-\lambda_j}\right] = 1 - F_2\left(-m_{ij}\right).$$

A similar proof can be presented for the RLPE model.

The latent variable s_{ij} is introduced to avoid working with Bernoulli type likelihoods, and this representation shows a latent linear structure producing equivalent models for the *LPE* or *RLPE* classes. Therefore, the *complete data likelihood function* for the skewed logistic IRT model is given by

$$L(\mathbf{a}, \mathbf{b}, \theta|s, y) = \prod_{i=1}^{n}\prod_{j=1}^{k} f_{\lambda}^{*}\left(s_{ij}\right)p\left(y_{ij}|s_{ij}\right), \tag{7}$$

where

$$p\left(y_{ij}|s_{ij}\right) = I\left(s_ij, yi_j\right) = I\left(s_{ij} > 0\right)I\left(y_{ij} = 1\right) + I\left(s_{ij} \leq 0\right)I\left(y_{ij} = 0\right), \; i = 1, \ldots, n,$$

$j = 1, \ldots, k$, and f_λ^* is the pdf of the distribution corresponding to the reflection of F_λ considered in Equation 6. That is, we use the pdf of the scobit distribution, if LPE model is assumed or the pdf of the burr-II distribution if the RLPE is assumed. In both cases, when $\lambda_j = 1$, the corresponding result for the $L(.)$ model follows, similar to the result presented in Albert (1992) for the probit IRT model. The result in Equation 6 can also be obtained by considering a latent linear structure for the skewed logistic IRT model, namely,

$$s_{ij} = m_{ij} + e_{ij}, \quad e_{ij} \sim F_\lambda^*(.); \tag{8}$$

that is, e_{ij} in the equation is an error term distributed as the standard reflection distribution considered in Equation 6. Note that the errors e_{ij} are independent and are "latent data" residuals (Albert & Chib, 1995). When considered $e_{ij} = s_{ij} - m_{ij}$, the error can be estimated using the residuals of the data and they can also be used for model checking. To understand how the observations y_{ij} change the distribution of these residuals, we consider the posterior distribution of e_{ij} conditional on a_j, b_j, θ_i and s_{ij}, that is, $e_{ij}^* = e_{ij} \big| a_j, b_j, \theta_i, y_{ij}, s_{ij}$. A similar analysis is presented in Fox (2004).

Note that in the skewed logistic IRT models, the parameters λ and a, b ,θ have quite different meaning regarding estimation. On one hand, λ is a vector of structural parameters associated with the choice of the link function. On the other hand, traditional IRT parameters of the Logistic IRT model a, b, θ are a vector of structural parameters inherent to the observed data and not depending on model choice (for a discussion, see, e.g., Taylor & Siqueira, 1996). By considering this fact, two scenarios can be considered. The first scenario is one in which λ and the traditional IRT parameters are estimated jointly; in the second scenario, only traditional parameters a, b, θ are allowed to vary and λ is fixed at its "true" value λ_0. As in Taylor and Siqueira (1996), we shall refer to these two scenarios as the unconditional and conditional ones for λ, respectively.

Inference under conditional scenario for λ is easier to be implemented from both maximum likelihood (ML) and Bayesian approaches because it corresponds to a particular logistic IRT model by considering a fixed value λ_0 for the parameter λ. However, conditions shall be imposed for the existence of the ML estimators and the posterior distribution of a, b, θ under improper uniform priors.

Regarding the unconditional approach for λ, computing the ML estimators using the versions of the likelihood functions given in Section 2 is not simple and is necessary to develop new conditions for the existence of the ML estimators. Additionally, it is important to study the propriety of the posterior distribution under improper uniform priors for b, θ and λ but considering proper priors for a. These aspects are being directed to another paper. In this article as well as in other articles dealing with IRT modeling, the existence of the ML estimators is assumed and proper priors are considered for the parameters in the model.

3.2. A Bayesian Approach

In this article, we adopt mainly the Bayesian view. Our point is made because several researchers demonstrated that accurate estimation of the item parameters in small samples can only be accomplished through a Bayesian approach (see, e.g., Swaminathan, Hambleton, Sireci, Xing, & Rizavi, 2003). Given the peculiarities of IRT models, ML totally relies on large sample theory, which even for a large number of examinees, can be complicated by the presence of incidental parameters. Researches using such an approach typically do separate estimation for item and ability parameters. However, there is no way to jointly evaluate estimates precision (Patz & Junker, 1999). Because of this, an expectation-maximization (EM) type algorithm as the one given by Bock and Aitkin (1981) is preferable. Such problems do not occur with the Bayesian approach in which, for a large number of examinees, the prior distribution has little effect on the posterior distribution (Sinharay & Johnson, 2003).

Prior specification. Prior specification is an important issue in Bayesian analysis. It is more important for small sample sizes where the posterior distribution represents more of a compromise between the observed data and previous personal opinion. For large sample sizes, it has less importance because the data typically dominate the posterior (information) distribution.

In the IRT literature, there seems to be consensus with respect to the prior for θ, that is, usually it is assumed that $\theta_i \sim N(0,1)$ for $i = 1, \ldots, n$, but different priors have been investigated for the traditional item parameters a_j and b_j (see Rupp, Dey, & Zumbo, 2004). Empirical evidence (see Patz & Junker, 1999, among others) seems to indicate the presence of posterior correlation between item parameters. However, it seems difficult to assign dependent priors for those parameters, being an especially hard task thinking about values for the correlations for such priors, even if a multivariate normal prior is specified. Hence, we prefer using independent and common priors for a, b, and λ and let such correlations be only data dependent. That is, the prior we consider can be written as

$$\pi(\theta, \mathbf{a}, \mathbf{b}, \lambda) = \prod_{i}^{n} \phi(\theta_i) \prod_{j}^{k} \pi_1(a_j)\pi_2(b_j)\pi_3(\lambda_j). \tag{9}$$

where $\phi(.)$ is the pdf of the standard normal distribution and $\pi_1(.), \pi_2(.), \pi_3(.)$ are the prior pdf for parameters $a_{j,}$ b_j, and λ_j, respectively.

Although some authors such as Albert (1992) and Fox and Glass (2001, 2003), use improper noninformative priors for the parameters $a_{j,}$ b_j, of the type $\pi_1(a_j)\pi_2(b_j) = I(a_j > 0)$, we prefer using informative priors on the discrimination parameters a_j because the existence of the joint posterior distribution is not guaranteed when an improper prior is used. Considering the results of Albert and Ghosh (1999), the distribution of the discrimination parameter must be proper to guarantee a proper joint posterior distribution.

Several informative distributions for a_j have been proposed in the literature. To mention just a few, (a) Bradlow, Wainer, and Wang (1999) and Albert (1992) use the $N\left(\mu_a, \sigma_a^2\right)$ with or without hyperparameter distributions specified for μ_a and σ_a^2, respectively; (b) Patz and Junker (1999) and Sinharay (2004) use the $LN\left(\mu_a, \sigma_a^2\right)$ with or without hyperparameter distributions specified for μ_a and σ_a^2, respectively, where LN(.) is the log-normal distribution; (c) Spiegelhalter, Thomas, Best, and Gilks (1996) and Sahu (2002) use the $HN\left(\mu_a, \sigma_a^2\right)$ where HN(.) is the half-normal distribution with known values for μ_a and σ_a^2; and, finally, (d) Swaminathan and Gifford (1985) use the IG(m, n), the inverted gamma distribution with (known) hyperparameter m and n. We consider in this article, the specifications in (b) above because $a_j > 0$ and also for conjugational reasons.

When independent informative priors are considered for the item parameters, it is usually assigned the $N\left(\mu_b, \sigma_b^2\right)$ for $b_j, j = 1, \ldots, k$. Moreover, in the common situation where little prior information is available about the difficulty parameter, one can chose σ_b^2 to be large. As is mentioned in Albert and Ghosh (1999) in the Logistic IRT model, this choice will have a modest effect on the posterior distribution for nonextreme data, and it will result in a proper posterior distribution when extreme data (all items are correct or all items are incorrect) are observed. Thus, vague priors can be used with the difficulty parameter.

In this article, we consider hyperparameters to be known. In more general situations, the prior structure needs to be enlarged so that hyper prior information can also be considered for those parameters. Following Sinharay (2004), we consider $\mu_a = 0$ and $\sigma_a^2 = 1$ for a_j, and then it is expected that $E\left(a_j\right) = e^{\mu_a + \sigma_a^2} = 1.649$ and $V\left(a_j\right) = \left(e^{\sigma^2} - 1\right)e^{2\mu_a + \sigma^2} = 4.671$; also, $\mu_b = 0$ and $\sigma_b^2 = 1$ for b_j.

The same prior distribution is considered for parameter λ_j. We consider the prior specified in the example 5.7 of Carlin and Louis (2000) in the context of binary regression using the power link. That is, $\lambda_j \sim$ gamma (0.25, 0.25). Thus, it is expected that $E(\lambda_j) = 1$, $V(\lambda_j) = 4$, with 2.5% and 97.5% quantiles being $q(0.025) = 1.055 \times 10^{-6}$ and $q(0.975) = 6.86647$, respectively.

MCMC Bayesian estimation. Considering the likelihood function in Equation 5 or Equation 7 and the general prior specification given in Equation 9, Bayesian estimation can be implemented by considering implementation of Markov chain Monte Carlo methods that make it simple to implement efficient sampling from the marginal posterior distributions.

In the first case, the specification by considering a hierarchical structure is implemented easily in WinBUGS (see Appendix) but not when the model is implemented using other programs. We point out that to implement a Bayesian estimation procedures involving a Bernoulli type likelihood can be complicated because the integrals involved to obtain the marginal posterior distributions with likelihood (Equation 5) are difficult. However, in the second case, by considering the latent structure in Equation 7, and the approach based on data augmentation that was introduced in the Section 3, the full conditionals for the skewed logistic IRT model and the Bayesian inference via MCMC follows without complications, similarly as reported in Albert (1992) when the normal ogive IRT model is implemented. Note that some of the full conditionals cannot be directly sampled from, requiring algorithms such as the Metropolis-Hastings. However, to implement the Bayesian approach in WinBUGS considering directly the likelihood function in Equation 7, it is necessary to have the cdf of the burr or scobit distribution which, to the best of our knowledge, is not yet implemented with the software.

In the remainder of this article, we develop a computational procedure for the skewed logistic IRT model based on the original likelihood function in Equation 5. Hierarchically, the full likelihood specification is given as follows:

$$y_{ij}|a_j, b_j, \theta_j, \lambda_j \sim \mathrm{Ber}\big(F_\lambda\big(a_j\big(\theta_i - b_j\big)\big)\big) \tag{10}$$

$$a_j \sim \pi_1(\mu_a, \sigma_b) \tag{11}$$

$$b_j \sim \pi_2\big(\mu_b, \sigma_b^2\big) \tag{12}$$

$$\lambda_j \sim \pi_3(m.n) \tag{13}$$

$$\theta_i \sim N(0, 1) \tag{14}$$

$$i = 1, \ldots, n;\ j = 1, \ldots, k.$$

Note that 1L and 2L can be obtained as particular cases of the hierarchical structure above. In addition, it is possible to introduce for the skewed logistic IRT model a guessing parameter c_j and then to substitute $F_\lambda\big(a_j\big(\theta_i - b_j\big)\big)$ by $c_j + \big(1 - c_j\big)F_\lambda\big(a_j\big(\theta_i - b_j\big)\big)$. In addition, a prior specification $c_j \sim \pi_4(r, s)$ should be considered. In this case, we can obtain 3LPE and 3RLPE models as generalizations of 3L model, which is obtained when $\lambda_j = 1$. When this is the case, the prior specified in Patz and Junker (1999) to c_j can be considered; that is, $c_j \sim$ Beta(5, 17), and then it is expected that $E(c_j) = 0.227$ and $V(c_j) = 0.0076$.

3.3. Model Comparison Criteria Using MCMC Outputs

A variety of methodologies exist to compare alternative Bayesian model fits but the principal criteria used in those works are the DIC proposed by Spiegelhalter et al. (2002) and the expected information criteria corresponding to Akaike (EAIC) and Schwarz or Bayesian (EBIC) as proposed in Carlin and Louis (2000) and Brooks (2002). The criteria are based on the posterior mean of the deviance: $E[D(a, b, \lambda, \theta)]$, where

$$D(\mathbf{a}, \mathbf{b}, \lambda, \theta) = -2\ln(p(y|\mathbf{a}, \mathbf{b}, \lambda, \theta)) = -2\sum_{i=1}^{n} \ln P\big(Y_{ij} = y_{ij}|\mathbf{a}, \mathbf{b}, \lambda, \theta\big),$$

which is also a measure of fit that can be approximated using the MCMC output, by considering $Dbar = \frac{1}{G}\sum_{i=1}^{G} D(a^g, b^g, \lambda^g, \theta^g)$, where the index g represents the gth realization of a total of G realizations and is the Bayesian deviance.

EAIC, EBIC, and DIC can be estimated using the MCMC output by considering

$$\widehat{\text{EAIC}} = Dbar + 2p,$$
$$\widehat{\text{EAIC}} = Dbar + p\log N,$$

and

$$\widehat{\text{DIC}} = Dbar + \rho\hat{D} = 2Dbar - Dhat,$$

respectively, where p is the number of parameters in the model; N is the total number, that is, $N = k \times n$ of observations; and ρD, the effective number of parameters, is defined as

$$\rho D = E[D(\mathbf{a}, \mathbf{b}, \lambda, \theta)] - D[E(\mathbf{a}), E(\mathbf{b}), E(\lambda), E(\theta)],$$

where $D[E(a), E(b), E(\lambda), E(\theta)]$ is the deviance of the posterior mean obtained when considering the mean values of the generated posterior means of the model parameters, which is estimated by

$$Dhat = D\left(\frac{1}{G}\sum_{i=1}^{G} a^g, \sum_{i=1}^{G} b^g, \frac{1}{G}\sum_{i=1}^{G} \lambda^g, \frac{1}{G}\sum_{i=1}^{G} \theta^g\right).$$

Given the comparison of two alternative models, the model that better fits a data set is the model with the smallest value of Dbar, DIC, EBIC, and EAIC. In EAIC and EBIC, $2p$ and $p\log N$ are fixed to penalize the posterior mean of the deviance whereas in IRT models, p is the number of parameters of the model (in the skewed logistic IRT model, $p = 3k + n$) and N is the total number of observations. Moreover, as there is no consensus in the use of the DIC (see discussion in Spiegelhalter, et al. 2002), the use of more than one criterion seems more appropriate to perform model comparison.

4. Illustration on Simulated Data

To evaluate the Bayesian estimation of the skewed logistic IRT models presented and evaluate the performance of model comparison criteria, we apply the estimation showed in Table 1 in Samejima (1999). The table shows, considering ML, the estimates of θ based on 32 response patterns of 5 dichotomous items

TABLE 1
Results Comparing the Skewed Logistic IRT Models With Logistic IRT Models

Models		p	*Dbar*	*Dhat*	DIC	EAIC	EBIC
Symmetric	1L	992	17,227.3	16,548.1	17,906.6	19,211.3	26,920.9
logistic IRT	2L	1,010	16,886	16,104.3	17,667.8	18,906	26,755.5
models	3L	1,028	17,021.1	16,506.7	17,535.4	19,077.1	27,066.5
Asymmetric	LPE	1,028	16,885.4	16,728.3	17,042.5	18,941.4	26,930.8
skewed logistic IRT models	RLPE	1,028	16,832.3	16,708.7	16,955.8	18,888.3	26,877.6

Note. DIC = deviance information criterion; EAIC = expected Akaike's information criterion; EBIC = expected Bayesian information criterion; IRT = item response theory; LPE = logistic positive exponent; RLPE = reflection logistic positive exponent.

following 2L, 2P, and LPE model with $\lambda = 2$, $a_j = 1$ for all items and $b = c$ $(-3.0, -1.5, 0.0, 1.5, 3.0)$, respectively.

Because ML estimation is not possible to extreme cases (response patterns 1, 1, 1, 1, 1 and 0, 0, 0, 0, 0), for the results to be comparable, we conduct Bayesian estimation for the other response patterns. In addition, as a and b are known parameters, we consider a vague prior for θ, that is $\theta_j \sim N(0, 1000)$, and use the posterior mean of θ as estimates to compare Bayesian and ML estimation. We consider a burn-in of 4,000 iterations and a chain of 2,000 iterations. Thin values for 2P, 2L, and LPE were, respectively, 1, 1, and 20.

By considering the sum of squares of the differences between the estimates, namely, $\sum_{i=1}^{30} \left(\theta_{i,\mathrm{ML}} - \theta_{i,B}\right)^2$, we found the minimum differences between the estimates of θ under ML and Bayesian approaches (the values were .044, .099, .221 for 2P, 2L, and LPE IRT models, respectively). It confirms the good performance of the Bayesian estimation approach, because for the minimum information priors considered, Bayesian and ML estimation are close when dealing with abilities estimation.

5. Application to Math Test

We illustrate the Bayesian approach developed in this article for skewed logistic IRT model with an application to real data set. We consider an analysis on the response pattern obtained by the application of a mathematical test to fourth-grade students of the rural Peruvian elementary schools. Item response vectors are available from authors on request and correspond to response of 974 students to 18 items qualified as binary responses (correct or incorrect). The scores present a mean of 8.27, a median of 8, and a standard deviation of 4.20. The skewness and kurtosis indexes are estimated as -0.075 and -0.836, respectively. The test presents a regular reliability index given by Cronbach's alpha of .83 and presents a mean proportion of items of .449.

TABLE 2
Item Parameters for Alternative IRT Models for Item 14, Item 2, and Item 11 in Math Data

Items	Models	Item Parameters			
		Discrimination, *a*	Difficulty, *b*	Guessing, *c*	Acceleration (deceleration), λ
Item 14	2L	1.003	−0.358		
	3L	1.065	−0.105	0.108	
	LPE	0.934	−1.214		2.087
	RLPE	4.559	−1.603		0.208
Item 2	2L	1.2777	−0.453		
	3L	1.607	−0.088	0.165	
	LPE	1.768	0.253		0.559
	RLPE	1.060	0.380		2.107
Item 11	2L	1.826	0.092		
	3L	2.065	0.220	0.067	
	LPE	1.680	0.039		2.060
	RLPE	2.070	0.023		1.046

Note. IRT = item response theory; LPE = logistic positive exponent; RLPE = reflection logistic positive exponent.

The mathematical test is formed with independent items corresponding to different tasks with different definitions. Given the latent ability θ, it is considered that the correct responses to the items are independent. Furthermore, the autocorrelations within individual responses seem to be low, which provides additional support for the assumption of local independence.

We present next a study on the fit of the parametric IRT models discussed earlier using the math data. Logistic IRT models with one parameter, two parameters, and three parameters are considered and denoted by 1L, 2L, and 3L, respectively. Moreover, we implement the Bayesian approach for the skewed logistic IRT models as discussed in Section 3.

Several criteria computed using the CODA package, including the ones proposed by Geweke (1992), were used to evaluate convergence, with good indication that such is the case.

DIC values shown in Table 1 seem to indicate that the skewed logistic IRT models (LPE and RLPE), improve any other proposed model including the corresponding symmetric ones (1L, 2L, and 3L). Hence, we expect that ICCs estimates are more precise with the skewed logistic IRT models. However, by considering EAIC, we found that RLPE is better, and by considering EBIC, we found that 2L is better. We prefer the interpretation of the parameters in the RLPE model because it is more consistent with the problem-solving process in some items, as is showed below.

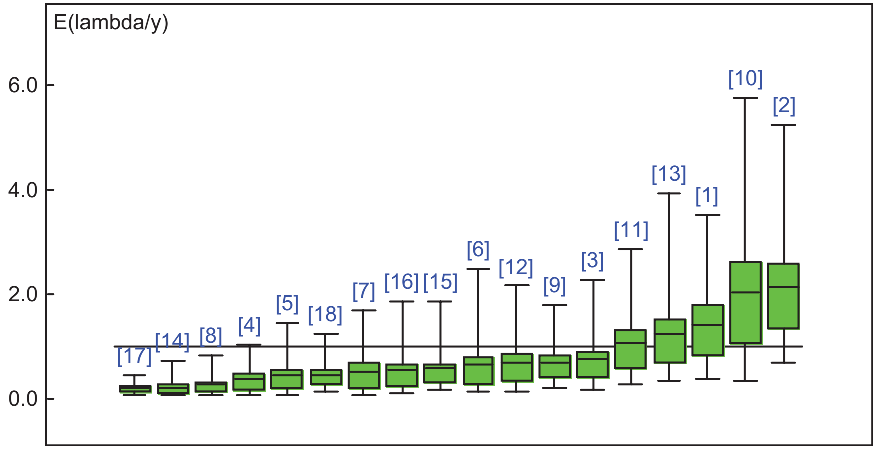

FIGURE 2. *Box plots for the λ parameters for the 18 items in the math data set under the reflection logistic positive exponent item response theory (RLPE IRT) model.*

By considering the RLPE model and according to Figure 2, we found that items 17, 14, 8, and 4 present significant asymmetries (i.e., the corresponding Highest Posterior density (HPD) intervals does not include $\lambda = 1$). Other items do no present significant asymmetries and can be modeled by considering 2L IRT models including Items 1, 10, and 2, which have posterior mean values of lambda greater than 1.

For an additional analysis, we choose 3 items: Item 14, clearly asymmetric; Item 2 with some degree of asymmetry; and Item 12 for which asymmetry is not evidenced. The items are presented in Figure 3.

Figure 4 depicts ICC fits considering RLPE IRT models in comparison with 2L, 3L, LPE IRT models to the Items 14, 2, and 11 with values of item parameters (posterior means) in Table 2. For these items, as shown in this figure, the ICCs under RLPE model and other alternative models are clearly different between them with relation to Item 14 in comparison with Item 2 where there is indication that the ICCs are similar (but not necessarily equal). However, for Item 11, the ICCs are practically equal.

Include Table 2 with Item 14, we prefer the interpretation given by the RLPE model because it is more consistent with the (inherent) mean of the process to solve it. For this item, estimated $\lambda = 0.2077$, hence different from 1, clearly showing differences among the ICCs of alternative models. That is, for ability values greater than -1.5, the probability of correct response to the item is greater with the RLPE model and for ability values lower than -1.5, the probability of correct response to the item is less with RLPE model and for ability values greater than 1.5, the probability of correct response to the item is greater with the RLPE model. Item 14, although easy, is not straightforward and can be considered a complex item in the sense of Samejima. It is a compound item and requires

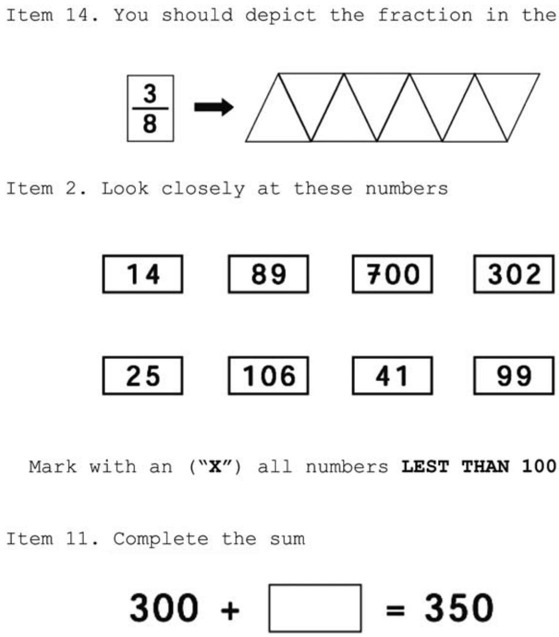

FIGURE 3. *Items 14, 2, and 11 for the math data set.*

a student to perform or successively pass each of many sequential subprocesses (parallel or serial, according to Samejima) to solve successfully the complete problem. However, this process is simple, and as the item also is an easy item, then students who solve this item should get greater credit in terms of probability of correct response.

Moreover, Item 14 corresponds to a limiting case and a type of heavyside step ICC function was obtained because the item has a large (estimated) discrimination parameter ($a = 4{,}559$). Thus, the ICC for Item 14 corresponds to a Guttman type ICC: at a very low ability level ($\theta < -1.5$), the probability of passing the item is practically zero, but at higher ability level, that is ($-1.5 < \theta < 1.9$), a small ability change is translated into a big change on the probability to pass the item. Finally, at high ability levels ($\theta > 1.9$), the probability of passing the item is practically 1.

Note that Item 2 also is an item requiring a student to perform or successively pass each of many sequential subprocesses to solve successfully the complete problem that can include series and parallel processes. It is a complex item with estimated $\lambda = 0.5588$, but by considering item estimated parameters, the ICCs are maybe similar among students with exception of the 3L model that consider higher value of the guessing parameter. This item cannot be considered totally complex in the sense of Samejima, because the value of lambda is not significantly different from 1.

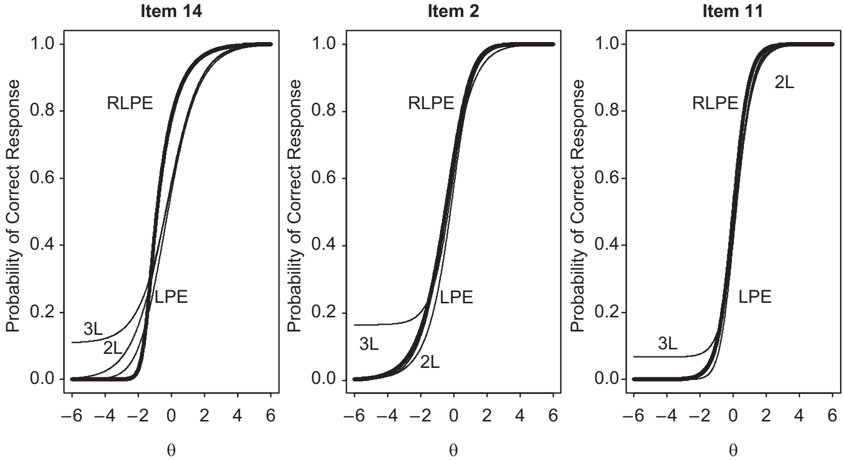

FIGURE 4. *Item characteristic curves (ICCs) for Items 14, 2, and 11 under the 3L, 2L, reflection logistic positive exponent (RLP) and reflection logistic positive exponent item response theory (LPE IRT) models in the math data.*

In addition, Item 11 is a typical item where sequential subprocesses are not required. For Item 11, all the ICCs are practically equal with estimated $\lambda = 1.046$ in the RLPE model, making its ICC indistinguishable from the ICC of a 2L model.

5. Final Discussion

In this article, we propose a new ICC to IRT, named Reflection of LPE or RLPE model, and implement a Bayesian approach to the already studied LPE model. The two models are named skewed logistic IRT models and is proved that one is the reflection of the other. This corresponds to the proposal of Samejima (1997a, 2000) for the fitting of asymmetrical IRT models and includes the symmetric logistic IRT model as a special case.

A data augmentation approach is proposed to implement Bayesian estimation using the MCMC methodology that can be implemented using the Metropolis-Hasting algorithm. Several model comparison criteria are used to compare the symmetrical and asymmetrical IRT models (DIC, EAIC, and EBIC). We also introduce latent residuals for the models and global discrepancy measures as the posterior sum of squares of the latent residuals. All these quantities show that the skewed logistic IRT model presents better fit than the usual logistic IRT model for the observed data. Moreover, to several items, there is clear indication that the penalization parameter λ, in the RLPE model, is different from one, indicated by the credibility interval, suggesting that asymmetric ICCs are more adequate. This parameter has a helpful interpretation in the context of the Peruvian mathematical test. Moreover, the penalization parameter is conceptually different from the guessing parameter in the three-parameter model, which is valid to the case of items with multiple choices (not considered in the example studied). Extensions of the model proposed will be subject of future developments, for example, to consider the inclusion of the new parameter in one-parameter and testlet models (Bradlow et al., 1999). Extensions to more general models such as multidimensional, hierarchical, and multilevel skewed logistic IRT models will also be the subject of future developments. Another possible direction is to conduct a sensitivity analysis of prior specification for the skewed logistic IRT model as the one conduced in Bazan, Bolfarine, and Leandro (2006).

Finally, the syntax in the appendix shows that the skew logistic IRT models can be easily implemented in applications.

Appendix
Program

We present next the program in WinBUGS used to implement the data augumentation approach described in the paper.

```
model{
for (i in 1:n) { for (j in 1:k ) {
  y[i,j]~dbern(p[i,j]) m[i,j]<-a[j]*(theta[i]-b[j])
  #LPE Model
  logist[i,j]<- exp(m[i,j])/(1+exp(m[i,j]))
  p[i,j]<-pow(logist[i,j],lambda[j])
#RLPE Model
  p[i,j]<-1-pow(1+exp(m[i,j]),-lambda[j]) }
              }
#abilities priors
  for (i in 1:n) { theta[i]~dnorm(0,1)}
#items priors
  for (j in 1:k) {
# usual priors
            b[j]~dnorm(0,1) a[j]~dlnorm(0,0.5)
#CARLIN and Lois (2000, Example 5.7)
lambda[j]~dgamma(0.25,0.25) }
  }
```

Acknowledgment

The authors acknowledge partial financial support from CNPq and Fapesp-Brasil and DAI-PUCP. They also acknowledge helpful comments and suggestions from the Editor and an anonymous referee, which substantially improved presentation.

References

Achen, C. H. (2002). Toward a new political methodology: Microfoundations and art. *Annual Review of Political Science*, *5*, 423–450.

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, *17*, 251–269.

Albert, J. H., & Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika*, *82*, 747–759.

Albert, J. H., & Ghosh, M. (1999). Item response modeling. In D. Dey, S. Ghosh, & B. K. Mallick (Eds.), *Generalized linear models: A Bayesian perspective* (pp. 173–193). New York, NY: Marcel-Dekker.

Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika*, *68*, 357–363.

Bazan, J., Bolfarine, H., & Branco, M. D. (2010). A framework for skew probit link in binary regression. *Communication in Statistics: Theory and Methods*, *39*, 678–697.

Bazan, J., Bolfarine, H., & Leandro, A. R. (2006). Sensitivity analysis of prior specification for the probit-normal IRT model: A empirical study. *Estadistica*, *58*, 170–171, 17–42.

Bazan, J., Branco, M. D., & Bolfarine, H. (2006). A skew item response model. *Bayesian Analysis*, *1*, 861–892.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters application of an EM algorithm. *Psychometrika*, *64*, 153–168.

Bradlow, E., Wainer, H., & Wang, X. (1999). A Bayesian random effect model for testlets. *Psychometrika*, *64*, 153–168.

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, BEST, CARLIN, and van der Linde (2002). *Journal of the Royal Statistical Society Series B*, *64*, 616–618.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis.* Boca Raton, FL: Chapman & Hall.

Chen, M. H., Dey, D. K., & Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, *94*, 1172–1186.

Czado, C., & Santner, T. J. (1992). The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference*, *33*, 213–231.

Fox, J.-P. (2004). Multilevel IRT model assessment. In L. A. van der Ark, M. A. Croon, & K. Sijtsma (Eds.), *New developments in categorical data analysis for the social and behavioral sciences* (pp. 227–252). London, England: Lawrence Erlbaum.

Fox, J.-P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 269–286.

Fox, J.-P., & Glas, C. A. W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*, *68*, 169–191.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 169–193). Oxford, NY: Oxford University Press.

Johnson, N., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (Vol. 1, 2nd ed.). Boston, MA: Houghton Mifflin.

Nagler, J. (1994). Scobit: An alternative estimator to logit and probit. *American Journal Political Science*, *38*, 230–255.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*, 146–178.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose-response curves. *Biometrika*, *32*, 761–768.

Rupp, A., Dey, D. K., & Zumbo, B. (2004). To Bayes or not to Bayes, from whether to when: applications of Bayesian methodology to item response modeling. *Structural Equations Modeling*, *11*, 424–451.

Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, *72*, 217–232.

Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika, 60*, 549–572.

Samejima, F. (1997a). Departure from normal assumptions: A promise for future psychometrics with substantive mathematical modeling. *Psychometrika*, *62*, 471–493.

Samejima, F. (1997b). *Ability estimates that order individuals with consistent philosophies*. Paper presented at the 1997 Meeting of the American Educational Research Association, Chicago, IL, March 24–28, 1997.

Samejima, F. (1999). *Usefulness of the logistic positive exponent family of models in educational measurement.* Paper presented at the 1999 Meeting of the American Educational Research Association, Montreal, Quebec, Canada, April, 19–23, 1999.

Samejima, F. (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristics curves. *Psychometrika*, *65*, 319–335.

Sinharay, S. (2004). Experiences with Markov chain Monte Carlo convergence assessment in two psychometric examples. *Journal of Educational and Behavioral Statistics*, *29*, 461–488.

Sinharay, S., & Johnson, M. S. (2003). *Simulation studies applying posterior predictive model checking for assessing fit of the common item response theory models* (ETS RR-03-26). Princeton, NJ: Educational Testing Service.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal Royal Statistical Society Series B*, *64*, 583–639.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1996). BUGS 0.5 examples (Vol. 1, Version i). Cambridge, England: University of Cambridge.

Swaminathan, H., & Gifford, J. (1985). Bayesian estimation in the two parameter logistic model. *Psychometrika*, *50*, 349–364.

Swaminathan, H., Hambleton, R., Sireci, S., Xing, D., & Rizavi, S. (2003). Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. *Applied Psychological Measurement*, *27*, 27–51.

Taylor, J., & Siqueira, A. (1996). The cost of adding parameters to a model. *Journal of the Royal Statistical Society*, *58*, 593–607.

Authors

HELENO BOLFARINE is a professor, Department of Statistics, Universidade de Sao Paulo, Rua do Matão, 1010 - Cidade Universitária - São Paulo - SP - Brasil - CEP 05508-090; hbolfar@ime.usp.br. His research interests include mixed and psychometric models.

JORGE LUIS BAZAN is a professor in the Department of Sciences at the Pontificia Universidad Católica del Perú, Av. Universitaria 1801, San Miguel, Lima-Perú; jlbazan@pucp.edu.pe. His research interests include mixed and psychometric models.

Manuscript submitted March 11, 2010
Revision received March 11, 2010
Accepted April 23, 2010