MEASUREMENT OF THE NONSENSE WORD FLUENCY: BAYESIAN APPROACH TO A ITEM RESPONSE MODEL WITH SPEEDEDNESS

Jorge Luis BAZÁN[1]
Luis VALDIVIESO[2]
Márcia D'Elia BRANCO[3]

▪ ABSTRACT: Speededness refers to the situation where the time limit on a standardized test does not allow substantial number of examinees to fully consider all items in a test, thus estimation using a common three parameter logistic item response model ($3PL$) can lead to contaminated estimates of the parameters in the model. This work proposes a simple Bayesian model to estimate both, personal and item parameters, from a test data with evidence of Speededness. The model is strongly related with a model proposed by Goegebeur, de Boeck, Wollack and Cohen (2008), but contrary to this a dependence structure in the personal parameters is not initially assumed. We conduct a case study to analyze a data set of Nonsense Word Fluency in Peruvian students, which presents high evidence of Speededness. Comparing the results on this data set of the $3PL$ and the proposed model we found, as expected, that some difficulty and discrimination parameters are overestimated under the $3PL$. Similar measures on the examinees abilities are discovered and new personal parameters: tolerance and propensity toward speededness are obtained considering the proposed model. Finally, future studies are suggested.

▪ KEYWORDS: Item response models; bayesian estimation; speededness; logistic model; personal latent variables; nonsense word fluency.

[1]Universidade de São Paulo - USP, Departamento de Matemática Aplicada e Estatística, CEP: 13560-970, São Carlos, SP, Brazil. E-mail: *jlbazan@icmc.usp.br*
[2]Pontificia Universidad Católica del Perú, Departamento de Ciencias, San Miguel, Lima, Peru. E-mail: *lvaldiv@pucp.edu.pe*
[3]Universidade de São Paulo - USP, Departamento de Estatística, CEP: 05311-970, São Paulo, SP, Brazil E-mail: *mbranco@ime.usp.br*

1 Introduction

Bayesian methods have become popular in social sciences, health sciences, education and psychology due to its flexibility in accommodating numerous models for different situations in data analysis. "Bayesian methods have become a viable alternative to traditional maximum likelihood-based estimation techniques and may be the only solution for more complex psychometric data structures" (RUPP *et al.*, 2004). One field in which Bayesian methods are naturally applied is Item Response Theory (IRT).

A unidimensional IRT model is a probabilistic model used to explain the response of n examinees to a set of k items by considering a unidimensional latent variable θ, associated with individual abilities, and a set of item parameters.

Consider a sequence of binary random variables $\{Y_{ij} : 1 \leq i \leq n\,; 1 \leq j \leq k\}$ associated with item responses and assume they are conditionally independent given θ_i, the latent variable related to the examinee's i ability. Here $Y_{ij} = 1$ if examinee i correctly answers item j and $Y_{ij} = 0$ otherwise.

A three parameter item response model assumes that the probability of a correct response is given by

$$p_{ij} = P[Y_{ij} = 1 \,|\, \theta_i\,,\, a_j\,,\, b_j,\;\; c_j] = c_j + (1 - c_j)F(m_{ij}). \qquad (1)$$

where c_j is a guessing parameter for item j, F is a cumulative distribution function (cdf) and $m_{ij} = a_j(\theta_i - b_j)$ is a latent linear predictor involving the discrimination item parameter a_j, the difficulty item parameter b_j and the latent variable θ_i associated with the examinee's i ability.

Two popular examples for F are the standard normal and the standard logistic distribution functions. For instance, the logistic three parameter item response model, referred here as *3PL*, is given by

$$p_{ij} = P[Y_{ij} = 1 \,|\, \theta_i\,,\, a_j\,,\, b_j,\;\; c_j] = c_j + (1 - c_j)\frac{exp(a_j(\theta_i - b_j))}{1 + exp(a_j(\theta_i - b_j))}. \qquad (2)$$

For fixed values of the item parameters a_j, b_j and c_j, we can draw the probability in (2) as a function of the ability θ_i. This is called the item characteristic curve (ICC). Figure 1 shows the ICC for the *3PL* with $a_j = 1.6, b_j = 0$ and $c_j = 0.2$.

Note that the item difficulty parameter b_j is the inflection point of the ICC curve on the horizontal axis; it shifts the curve from left to right as the item becomes more and more difficult. Also b_j is in the same scale of θ_i, with $-\infty < b_j < \infty$. When $\theta_i = b_j$ we have that $p_{ij} = (1 + c_j)/2$ and this is the point where the slope is maximum. Higher values of b_j indicate that greater skill is required to correctly answer the item.

The item discrimination parameter a_j is assumed to be positive, since in our context the ICC curve should be increasing. This parameter is associated with the slope of the curve. Higher values of a_j indicate a more marked change in the probability of success for a fixed variation on the ability.

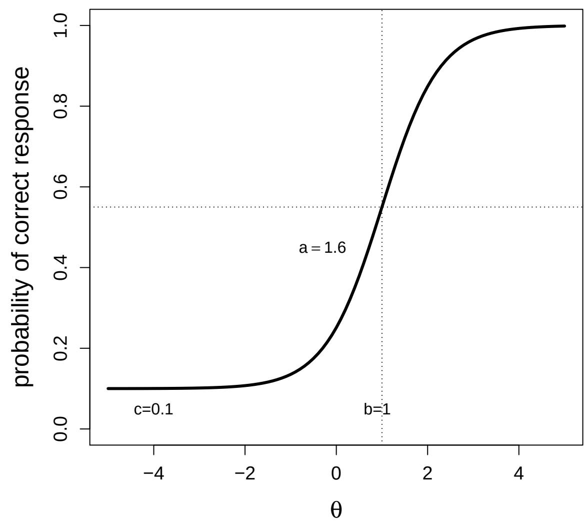

Figure 1 - Characteristic Curve for the *3PL* with $a = 1.6, b = 1$ and $c = 0.1$.

The lower asymptote is given by the guessing parameter c_j. It is the probability of a correct response for an examinee with infinitely low skill. Theoretically it could range from 0 to 1, but more realistically it is not more than 0.3.

Note that the number of parameters involved in the model is $3k + n$, which makes it difficult to obtain the maximum likelihood estimates. Classical solutions for this problem consider a probability distribution for the latent variables θ_i and use a two step procedure (BAKER and KIM, 2004). In the first step, the likelihood function is integrated over the latent variable and this marginal likelihood is maximized to obtain estimates for the items parameters. These parameters are then fixed to maximize the original likelihood and obtain the θ_i estimates. Bock and Aitikin (1981) propose a pseudo-EM algorithm to implement the first step. Limitations of this methodology are discussed in Patz and Junker (1999) and Sahu (2002).

In this paper we will focus on IRT speededness models. These are IRT models in which examinees have a time limit. Popular unidimensional IRT models assume that the probability that an examinee gives a correct response to an item depends only on the examinee´s proficiency and the characteristic of the item. In addition, there is an assumption that the examinee has sufficient time to answer all items in the test. However, "A test is speeded when some portion of the test taking population does not have sufficient time to attempt every item in the test within the allocated time" (BEJAR, 1985). So, when speededness exists in a test not designed to take it into account and the speed is not modeled as examinee's proficiency, the IRT assumptions are violated as examinees may fail to give correct responses not because of limited proficiency, but because of limited time. In such a case new models considering other personal variables together with the examinee's proficiency and the characteristics of the items need to be addressed. We will focus in this paper on a model when speededness is observed, under the Bayesian framework.

In the next section, we discuss the fully Bayesian approach for IRT using MCMC. In section 3, the concept of speededness is presented and some approaches to model this latent characteristic are discussed. A simple Bayesian model to estimate both, personal and item parameters, from a data with evidence of speededness is presented in section 4. This model is strongly related with the speeded item response model with gradual process change proposed by Wollack and Cohen (2005) and Goegebeur, de Boeck, Wollack and Cohen (2008) but differs from this in the sense that no dependence structure is initially assumed in the formulation of the model. Because of this, we will name our model an *alternative 3PL speeded* model. In section 5, we conduct a comparative study to analyze the 3PL and the alternative 3PL speeded model with a data set of Nonsense Word Fluency, which presents high evidence of Speededness. The data was obtained from the Early Grade Reading Assessment (EGRA) instrument (RTI-FDA, 2008) designed to measure pre-reading skills in peruvian students. Finally, conclusions and future works are presented in section 6.

2 Bayesian approach in IRT

As pointed out by Rupp *et* al. (2004), a fully Bayesian analysis is taken when prior distributions are specified for all unknown quantities. In the classical IRT literature empirical Bayes estimators are, as explained before, used to estimate the latent parameters after obtaining the maximum likelihood item parameter estimators. This approach is taken in traditional commercial programs like *BILOG* or *MULTILOG*. In this chapter we will restrict our analysis to a fully Bayesian approach using Markov Chain Montecarlo (MCMC) methods where, instead of using *plug-in* estimators and a two step procedure, all parameter will be estimated together using the complete posterior distribution. Maximum Likelihood estimation to some common IRT models proposed in the literature are available for free as a package in R program (http://cran.r-project.org). See for example the ltm package of Rizopoulos (2006), the eRm package (MAIR and HATZINGER, 2007) and the lme4 package (DE BOECK *et al.*, 2011). On the other hand, only the MCMC package (GELMAN and HILL, 2007) and the pscl package (JACKMAN, 2008) allow to implement some common IRT models under the bayesian approach.

A seminal paper to this respect is Albert (1992) where a data augmentation procedure was suggested for the two parameters probit model [i.e, with $c_j = 0$ and a normal cdf in (1)]. Béguin and Glas (2001) generalized this approach for the three parameter probit and a multidimensional model, while Fox and Glas (2001) extended this approach for multilevel IRT models. Further details are given in Fox

(2010). In addition, alternative models with asymmetric links have been proposed by Bazán, Branco and Bolfarine (2006) and Bolfarine and Bazán (2010). While the first authors introduced a skew-probit IRT model using the cdf of a skew-normal distribution (AZZALINI, 1985), the last considered a skew version of the logistic model. In both cases a new set of item parameters to control the skewness of the distribution is introduced. The Bayesian inference for these models were performed using a data augmentation approach similar to Albert (1992).

In general, by considering the likelihood function and the correspondent prior specification, a joint posterior distribution for these IRT models can be easily fitted using MCMC, for instance throughout *WiNBUGS/ OpenBUGS* (CURTIS, 2010) or *SAS* (STONE and ZHU, 2015). MCMC algorithms can be quite sophisticated, their proper use require careful attention to several aspects of the implementation and demand a heavy computational knowledge (KIM and BOLT, 2007). This may hinder the large-scale use of this methodology by more applied researchers. The use of the BUGS software (www.mrc-bsu.cam.ac.uk/bugs) offers many advantages, including its relative ease of implementation, velocity and availability as a free software. Moreover, the use of BUGS software to estimate IRT models allows to change existing code in order to fit new variations of the current models, which cannot be fitted in the main classical software packages (IRT PRO, CAI *et al.*, 2011). For more details see Curtis (2010) and Bazán (2012).

3 Speededness in IRT

In the usual context of a unidimensional IRT the probability that an examinee gives a correct response to an item depends only on the examinee´s proficiency and the characteristic of the item. As a consequence the theory assumes that the examinee has sufficient time to answer all items in the test, ignoring speededness effects. According to Lu and Sireci (2007) speededness refers to the situation where the time limit does not allow substantial number of examinees to fully consider all items on a standardized test.

When speededness exists in a test not designed to take it into account and the speed is not modeled as examinee's proficiency, the IRT assumptions are violated and the examinees may fail to give correct responses not because of limited proficiency, but because of limited time. Hence, undetected speeded responses allow a time factor to contaminate the ability estimates. Furthermore, since examinees who are running out of time will often either hurry through the latter stages of a test or omit or randomly complete items towards the end of the test, these items tend to appear harder than they do when they are administered under nonspeeded conditions (OSHIMA, 1994).

As indicated by Yamamoto and Everson (1997), traditional methods of assessing test speededness are limited to the analysis of missing response distributions, especially at the end of a test. However, analysis of missing responses is inadequate in evaluating the speededness of multiple choice tests when the test score or *raw score* is a function of the total number of correct responses. A revision of these methods can be found in Lu and Sireci (2007).

In the past few years, several models that account speededness effects have been developed. Yamamoto and Everson (1997) have proposed a hybrid model which assumes that a two parameter logistic IRT model is appropriate throughout most of the test up to a point at which examinees switch to response randomly. Bolt, Cohen and Wollack (2002) have proposed a two class mixture Rasch model with ordinal constraints. These constraints are designed to distinguish a class with no speededness effects from a class whose responses are affected by speededness. More precisely, for items early in the test, the item difficulty parameters are constrained to be equal in both classes; however, the item difficulty parameters of end-of- test items in the speeded class are constrained to be larger than the respective item difficulty parameters in the nonspeeded class. Using a more general IRT model, Goegebeur *e*t al. (2008) have proposed a speeded item response model with gradual process change. This model was first formulated by Wollack and Cohen (2005) as a model to simulate speededness data and can be regarded as a multidimensional IRT model that extends the three parameter logistic model by considering three personal latent variables to explain the behavior of the examinees in the test under a speeded situation. We refer to this model as a *3PL speeded* model.

An alternative approach to modeling speededness has been considered by Van der Linden (2007) and Klein Entink, Fox and Van der Linden (2009). They assume a separate model for response time in order to measure speed of working besides an IRT model for measuring accuracy. At a higher level their models postulate a correlation between working speed and accuracy.

4 Two 3PL speeded IRT model

4.1 The *3PL speeded* IRT model

The *3PL speeded* IRT model with gradual process change (GOEGEBEUR *et al.*, 2008) extends the *3PL* IRT model by taking into account speededness effects. The model is formulated as:

$$Y_{ij}|\boldsymbol{\beta}, \boldsymbol{\Theta} \sim Bernoulli\,(\,p_{ij}), \;\; i = 1, \ldots, n, \;\; j = 1, \ldots, k. \tag{3}$$

where $\boldsymbol{\beta}$ is a $k \times 3$ matrix containing the vector item parameters: $\boldsymbol{a} = [a_1, a_2, \ldots, a_k]^{'}$, $\boldsymbol{b} = [b_1, b_2, \ldots, b_k]^{'}$ and $\boldsymbol{c} = [c_1, c_2, \ldots, c_k]^{'}$, $\boldsymbol{\Theta}$ is a $n \times 3$ matrix containing the vector personal latent variables or personal parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]^{'}$,$\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_n]^{'}$ and $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]^{'}$ and $\boldsymbol{y} = [y_{ij}]$ is the data matrix containing the observed dichotomous responses from (3) of the n examinees to the k items. Specifically, Y_{ij} is the dichotomous response corresponding to the i-th individual to the j-th item, the jth row of $\boldsymbol{\beta}$, $\boldsymbol{\beta_j} = (a_j, b_j, c_j)$, is the vector of item parameter that includes, respectively, the *discrimination*, *difficulty* and *guessing* parameters of item j and the jth row of $\boldsymbol{\Theta}$, $\boldsymbol{\Theta_i} = (\theta_i,\ \eta_i,\ \lambda_i)$, is the vector of personal parameter that includes, respectively, the *ability*, *tolerance towards speededness* and *propensity to guessing under speededness* of examinee i in answering the test with the k items. In addition, it is assumed that p_{ij}, the *probability of a correct answer* for examinee $i \in \{1, 2, \ldots, n\}$ to item $j \in \{1, 2, \ldots, k\}$, is given by

$$p_{ij} = P[Y_{ij} = 1 \,|\, \theta_i \,, \eta_i \,, \lambda_i \,, a_j \,, b_j ,\ \ c_j] = c_j + (1 - c_j)G(m_{ij}), \tag{4}$$

where

$$G(m_{ij}) = F(m_{ij})P_j(\eta_i, \lambda_i),$$
$$P_j(\eta_i, \lambda_i) = min\Big\{1, r_j(\eta_i, \lambda_i)\Big\}, \tag{5}$$

and

$$r_j(\eta_i, \lambda_i) = \Big[1 - (\frac{j}{k} - \eta_i)\Big]^{\lambda_i},$$

being $m_{ij} = a_j(\theta_i - b_j)$ a linear function of θ_i.

Note that in comparison with the *3PL* IRT model, two additional personal parameters associated with speededness effects, η_i and λ_i, are introduced in the model. We have called them, respectively, the *tolerance towards speededness* and the *propensity to guessing under speededness*. Observe that $\eta_i \in [0, 1]$ identifies the point j/k, expressed as a fraction of the number of items, where examinee i first experiences an effect due to speededness, while $\lambda_i > 0$ controls the rate of decrease towards a guessing situation.

The rationality behind model (3) is as follows. When examinee i encounters item j, the examinee answers according to a *3PL* IRT model or a random guessing process, with probabilities $P_j(\eta_i, \lambda_i)$ and $1 - P_j(\eta_i, \lambda_i)$ respectively. Under the problem solving process the examinee knows the answer with probability $F(m_{ij})$; if ignorant, the examinee guesses at random.

Note that $P_j(\eta_i, \lambda_i)$ in (5) can be regarded as a *penalizing factor due to speededness*. When speededness is not present in item j, i.e., if examinee i has a relatively high tolerance $\eta_i \geq \frac{j}{k}$ towards speededness, then $r_j(\eta_i, \lambda_i) > 1$ and $P_j(\eta_i, \lambda_i) = 1$. Consequently $F(m_{ij})$ is not decreasing. On the other hand, if examinee i has a relatively low tolerance $\eta_i < \frac{j}{k}$ towards speededness, then $r_j(\eta_i, \lambda_i) < 1$ and consequently $P_j(\eta_i, \lambda_i) < 1$ decreases the probability $F(m_{ij})$. Moreover this factor can be magnified by having a high propensity λ_i to guessing under speededness.

The parameter η_i can be considered for the personal latent variable associated with the tolerance to speededness in contrast with the relative position j/k of the item in the test. This relation is similar to the one between ability and difficulty. Both η_i and j/k should be on the same scale (between 0 and 1), despite the fact that η_i is a random variable and j/k is fixed.

High values of λ_i are associated with the increase of the penalization factor $P_j(\eta_i, \lambda_i)$ and therefore the decrease of the probability of correct response. However, note that the η_i parameter seems to be more relevant than the λ_i parameter because it can be interpreted in any scenario, while the second only under a speededness situation. Thus the λ_i parameter express a propensity to guessing which does not necessarily affect the probability of a correct response.

To complete the formulation Goegebeur *et al* (2008) considered a personal vector parameter $\mathbf{\Theta}_i$ with a joint distribution function:

$$G(\theta, \eta, \lambda) = C(G_1(\theta), G_2(\eta), G_3(\lambda)). \tag{6}$$

where C is a Gaussian copula and $G_1(.)$, $G_2(.)$ and $G_3(.)$ are respectively the marginal cumulative distribution functions of the personal parameters θ_i, η_i, and λ_i.

Thus, combinating (3-5) and (6) we have finally the *3PL speeded* IRT model proposed by Goegebeur *et al* (2008).

To estimate the parameters in this model and recalling that item responses are conditionally independent given Θ_i and there is independence among examinees, Goegebeur *et al* (2008) used the marginal maximum likelihood methodology. This consists on first maximizing the marginal likelihood function

$$L = \int_{-\infty}^{\infty} \int_0^1 \int_0^{\infty} L(\boldsymbol{\beta}, \mathbf{\Theta}|\boldsymbol{y}) dG(\theta_i, \eta_i, \lambda_i), \tag{7}$$

where

$$L(\boldsymbol{\beta}, \mathbf{\Theta}|\boldsymbol{y}) = \prod_{i=1}^{n} \prod_{j=1}^{k} p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}}, \tag{8}$$

with respect to the item and copula parameters and the using these estimates for a second stage to obtain an empirical Bayes estimation of the personal parameters.

The choice of a Gaussian copula was justified by the possibility to regard the model as a linear mixed model and then use the NLMIXED SAS procedure in the estimation process.

4.2 The *alternative 3PL speeded* IRT model

A disadvantage of the *3PL speeded* IRT is that a consideration of a dependence structure among the latent personal variables makes it difficult not only the multidimensional evaluation of (7) but also involves an increase in the number of parameters (the copula parameters). To overcome these issues we propose in this paper an alternative model by starting with (3-5), but assuming that the personal latent vector for any examinee i, $\mathbf{\Theta_i} = (\theta_i, \eta_i, \lambda_i)$, has initially independent components; that is we assume for $\boldsymbol{\theta_i}$ the joint distribution:

$$g(\mathbf{\Theta}_i) = g(\theta_i, \eta_i, \lambda_i) = g_1(\theta) \times g_2(\eta) \times g_3(\lambda) \tag{9}$$

where $g_1(.)$, $g_2(.)$ and $g_3(.)$ are continuous probability distribution functions defined in the correspondent parametric space of each personal latent variable. We will refer to the model (3-5) and (9) as the *alternative 3PL speeded* IRT model.

Note that under this Bayesian formulation the natural correlations between the personal latent parameters will depend on the data and will appear in the posterior distribution. This argument is similar to the case of assuming prior independence on the item parameter where empirical evidence (see Patz and Junker, 1999, among others) suggests the presence of a posterior correlation between item parameters.

4.3 Bayesian estimation of the *alternative 3PL speeded* IRT model

The *alternative 3PL speeded* IRT model assumes, as usually considered in the Bayesian literature, independent priors for all the parameters in the model. The following item parameter prior distributions are commonly specified in IRT modelling:

$$a_j \sim LN\,(0, 0.25) \text{ or } TN(0, 0.5)I(a > 0),\ b_j \sim Normal\,(0, 1) \text{ and } c_j \sim Beta\,(5, 17). \tag{10}$$

where LN and TN denote the log-normal and truncated normal distributions, respectively. The $LN(0, 0.25)$ (Patz and Junker, 1999) and $TN(0, 0.5)I(a > 0)$ (BAZÁN *et al.*, 2006) are non-negative distributions with mean-standard deviation pair $(1.13, 0.60)$ and $(1.11, 0.61)$, respectively. On the other hand, the $Beta(5, 17)$ distribution (FU *et al.*, 2009) has a mean-standard deviation pair $(0.23, 0.09)$. The distribution for b is justified using a similar argument given to choose the normal for θ. See, for instance, Albert and Ghosh (2000) and Sinharay *et al.*, (2006). Instead of using the usual parametrization we will opt in this work for the parametrization $m_{ij} = a_j\theta_i - b^*_j$ where the item difficulty parameter is easily recovery using $b_j = b^*_j/a_j$. In this case a prior specification for b^*_j similar to b_j is adopted. This is a common practice in Bayesian Inference (see, for example, Sahu (2002)).

To complete the specification of the personal parameters we will set the following priors:

$$g_1(.) : \theta_i \sim N(0, 1),\ \ g_2(.) : \eta_i \sim Beta\,(\alpha, \beta) \quad \text{and} \quad g_3(.) : \lambda_i \sim LN\,(\mu_\lambda, \sigma^2_\lambda). \tag{11}$$

As frequently considered in the literature (RUPP *et al.*, 2004), we take $g_1(.) : \theta_i \sim N(0, 1)$, $i = 1, \ldots, n$. This assumption establishes that it is believed that the latent variables are well behaved and abilities conform a random sample from this distribution. This additionally establishes a metric for the abilities estimates.

Specifications for hyperparameters of $g_2(.) : Beta\,(\alpha, \beta)$ can be made by considering information about the point where an examinee is more likely to experience a speededness situation. This can be obtained from judges or experts by asking them about the mean (μ_η) and standard deviation (σ_η) of η_i. For example, a choice can be $(\mu_\eta = 0.67, \sigma_\eta = 0.18)$ which correspond to a $Beta(4, 2)$ distribution, which means a moderate tolerance towards speededness. Other possibilities could

be a $Beta(2,9)$ or a $Beta(9,2)$ distribution with, respectively, $(\mu_\eta = 0.18, \sigma_\eta = 0.01)$ and $(\mu_\eta = 0.8, \sigma_\eta = 0.01)$. In the former case the tolerance towards speededness is low while in the second case high. Situations where no information is given can be modeled with a $Beta(1,1)$ distribution, or equivalently, a uniform distribution on the interval $(0,1)$.
For the λ personal parameter of propensity, Goegebeur *et al.* (2008) considered $g_3(.)$: $LN(\mu_\lambda, \sigma_\lambda)$ with $\mu_\lambda = -3.604$ and $\sigma_\lambda = 2.771$. After simulating 1000 samples from this distribution we found a maximum value of 17140.48 and a minimum of 2.497×10^{-8}. This scale is too broad for any meaningful interpretation of the distances between the values, so in order to shrink the scale, we set the hyperparameters $\mu_\lambda = 0$ and $\sigma_\lambda = 1$, which yields a mean of 1.65 and standard deviation of 2.16.

Since the posterior density in the likelihood function (8) and the priors in (10) and (11) can not be fully obtained in closed form we will use a Markov chain Monte Carlo (MCMC) approach to simulate parameter values and obtain parameter estimates. Given the hierarchical structure of our model, it is not difficult to implement such a process throughout several packages as OpenBugs, WinBugs, JAGS, Rstan or SAS. In this paper the model was fit using WinBUGS and the codes are available under requirement.

4.4 Simulation study

To assess the performance of our Bayesian approach we have simulated data set of the proposed model under four scenarios: $(n,k) = (500,20),(1000,20),(500,40),(1000,40)$. Values of $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ were generated from a $LogNormal(0,0.25)$, $Normal(0,1)$ and $Beta(12.5,37.5)$ distributions, respectively, and values of the personal latent variables $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ from a $Normal(0,1)$, $Beta(4,2)$ and $LogNormal(0,0.8^2)$ distributions, respectively. A total of R = 10 replications, i.e., ten sets of answers for the individuals to the items, were simulated from the proposed model under the four scenarios. All parameters were estimated using the proposed MCMC algorithm using an effective sample size of 1000. More precisely, 20000 iterations were taken, the first 10000 discarded, as a burn-in period, and a lag of size 10 was considered to eliminate potential problems due to autocorrelation. The convergence of the MCMC chains was monitored using trace plots, auto-correlation (ACF) plots and the criteria proposed by Geweke (1993). These provided to us in all cases with strong indication of chain convergence. The following priors were considered: $a_j \sim N(1,0.5)I(0,)$, $b_j \sim Normal(0,1)$, $c_j \sim Beta(5,17)$, $\theta_i \sim Normal(0,1)$, $\eta_i \sim Beta(1,1)$ and $\lambda_i \sim LogNormal(0,0.8)$.

We report in Table 1 the mean and standard deviation (SD) of the fitted and simulated values, together with the Bias and Mean Square Error (MSE) of the parameters for each set of parameters. For example, in order to measure the accuracy of the estimates the MSE is calculated as: $MSE(\boldsymbol{\kappa}) = \sum_{r=1}^{R} \frac{(\hat{\boldsymbol{\kappa}}_r - \boldsymbol{\kappa})^2}{R}$, where $\hat{\boldsymbol{\kappa}}_r$ is the estimated value of the parameter $\boldsymbol{\kappa}$, which can be $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{\theta}, \boldsymbol{\eta}$ or $\boldsymbol{\lambda}$.

By observing the results in Table 1, we obtain relatively small RMSE values for the parameters, except for the λ parameter. In performing the analysis of the bias, this parameter shows some unwanted values (zero distant). In general, as the sample size and the number of items increases, the performance of the estimation method improves. Thus, the results show that our Bayesian approach performs well in order to recover the true parameters, specially when higher values of sample size and of the number of items are considered.

5 A case study with evidence of speededness: the NWF data set

The Nonsense Word Fluency (NWF) data comes from a pilot study to adapt the Snapshot of School Management Effectiveness (SSME) and Early Grade Reading Assessment (EGRA) instruments to the Peruvian reality (RTI-FDA, 2008). The EGRA instrument includes a battery of test that are designed for the initial years in primary school and provides a quick information regarding school management and pre-reading skills. The NWF data was obtained after evaluating 512 eight years old students who were able to correctly read each of the 50 nonsense words in the NWF test within a time limit of 60 seconds. Figure 2 shows a subset of this test.

pul	quibe	ino	mise	jud	5
udo	zel	bedi	cur	miz	10
lline	rite	duso	jafi	fica	15
⋮	⋮	⋮	⋮	⋮	⋮
quira	cuto	dofu	afo	duba	50

Figure 2 - Subset of the Nonsense Word Fluency data in the EGRA instrument.

Figure 3 presents the percentage of students who were able to correctly read each of the 50 nonsense words. We see that this percentage decreases at the end of the test giving the impression that the last words are the hardest and suggesting the presence of speededness.

In figure 3 we considered non-reached words as incorrectly read. This certainly could introduce a bias that is needed to be addressed. An alternative for a more reliable speed measure than the percentage of success was proposed by Stafford (1971) and discussed in Lu and Sirecci (2007). Under the assumption that examinees keep no track of time during the test, which seems to be the case in our application, Stafford suggests to use the Speededness Quotation (SQ) index

$$SQ = \frac{\sum_{i=1}^{n} U_i}{\sum_{i=1}^{n} W_i + \sum_{i=1}^{n} U_i},$$

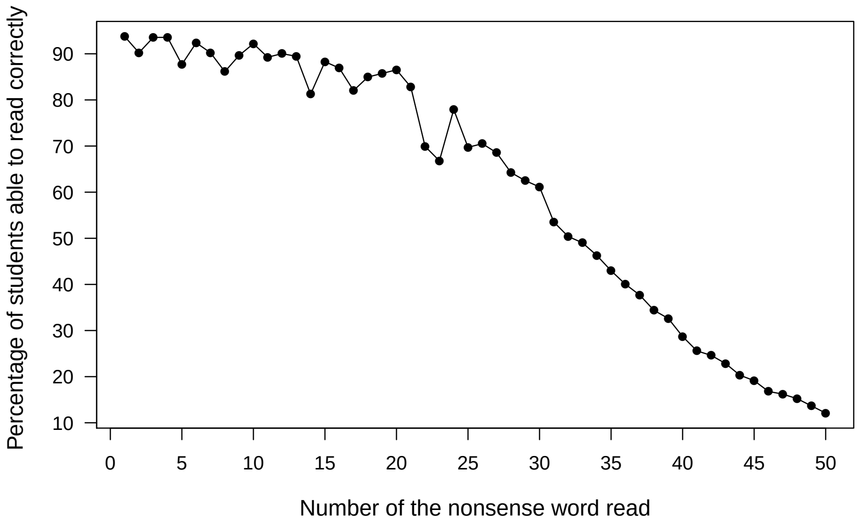

Figure 3 - Percentage of students who were able to correctly read each of the Nonsense Words in the NWF data.

where n is the number of examinees, U_i is the number of non-reached items by examinee i and W_i is the number of the incorrectly answered or omitted items by examinee i. Indices closer to 1 give evidence of speededness.

In the EGRA application we obtained a Speededness Quotation index of 0.9081, that is, data with evidence of speededness which supports the use *an alternative 3PL speeded* IRT model instead of the usual *3PL* IRT model or the classical raw score model to measure NWF.

To evaluate the *3PL* IRT model and the *alternative 3PL speeded* IRT models with this data we will consider the following prior distributions: $\theta_i \sim Normal(0, 1)$, $\eta_i \sim Beta(1, 1)$, $\lambda_i \sim LN(0, 1)$, $a_j \sim TN(1, 0.5)I(0,)$, $bs_j \sim Normal(0, 2)$ and $c_j \sim Beta(5, 17)$, where $b_i = bs_i/a_i$ is the difficulty parameter and bs_i is an intercept parameter. For both models we will assume a logistic cumulative distribution function in (4), which means that the *3PL* IRT model becomes the 3PL.

Starting with a burn-in of 5000 iterations and a thinning of 20, a sample of size 2000 was obtained. Several criteria using the CODA package, were computed for the convergence analysis.

Table 2 shows the fit comparison using the Dbar and DIC criteria (SPIEGELHALTER *et al.*, 2002). As seen, the *alternative 3PL speeded* model shows a better fit than the *3PL* for the NWF data. This confirms the evidence given by the SQ index.

Table 3 summarizes the estimation results for the personal latent variables under the *3PL* and the *alternative 3PL speeded* IRT model. We have also considered the raw scores to measure Nonsense Word Fluency. Correlations between personal latent variables are also included. Figure 4 shows the scatter plot between classical raw scores and ability estimations under the *3PL* and the *alternative 3PL speeded* IRT model. Variables were standardized to facilitate the comparison.

Observe from Table 3 and Figure 4 that the abilities under the *3PL* and the classical raw scores are highly and positively correlated ($r = 0.996$), so they could be interchangeable. On the other hand, abilities under *3PL* have a relatively lower correlation ($r = 0.725$) with abilities under the alternative *alternative 3PL speeded* model. This can be understood by observing that a group of examinees with low abilities under the alternative 3PL speeded model obtained correspondent higher measures of abilities under the *3PL* model. We consider this group as a group with overestimated abilities.

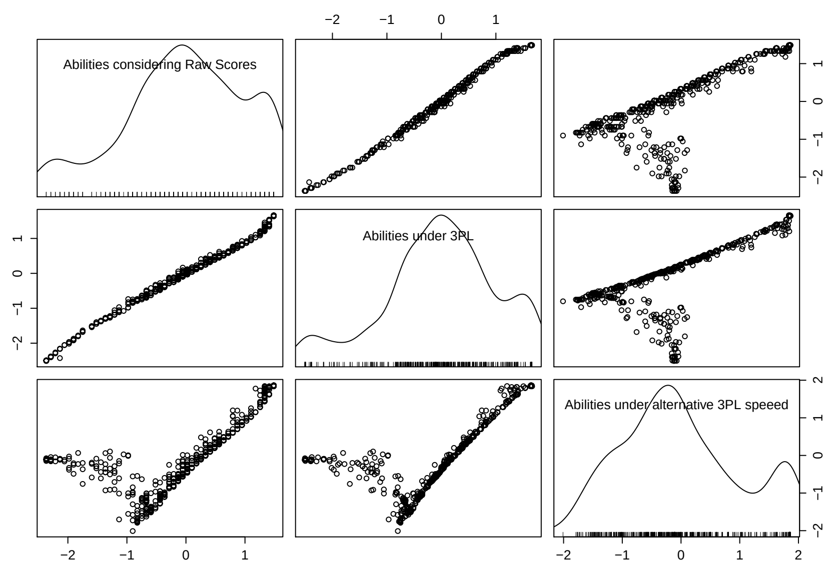

Figure 4 - Scatter plot for abilities estimates under the *3PL*, the *alternative 3PL speeded* IRT model and *raw scores* for Peruvian students in NWF data. Variables were standardized to facilitate the comparison.

As expected, Table 3 and Figure 5 show that considering the *alternative 3PL speeded* model θ and η are positively correlated ($r = 0.446$), which means that a low tolerance towards speededness is mainly associated with a low ability or NWF measure. On the other hand, the tolerance towards speededness and the propensity to guessing under speededness are highly and negatively correlated ($r = -0.712$).

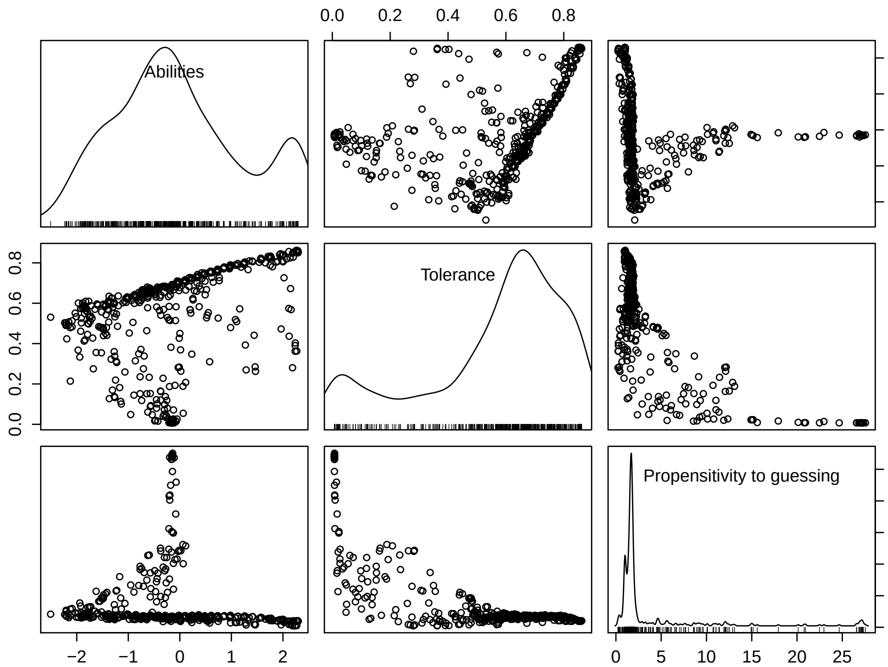

Figure 5 - Scatter plot for the personal latent variables estimates under the *alternative 3PL speeded* IRT model for Peruvian students in NWF data.

Moreover, a low negative correlation but significant, is observed between ability and propensity ($r = -0.136$) indicating that for some examinees high propensity to guess is associated with low ability or NWF measure. The importance of this type of results is that examinees can be best characterized in relation to personal latent variables which is important in the assessment of Nonsense Word Fluency, an important predictor of Reading Proficiency (see Fien *et al.*, 2008).

For a better understanding of the personal latent variables to the group identified in Figure 4, we have further studied some characteristics of this group with overestimated abilities under the *3PL* model. The results are shown in Table 4.

Table 1 - Perfomance of the Bayesian approach for the *alternative 3PL speeded* model in the simulation study

		$n = 500,\ \ k = 20$				$n = 1000,\ \ k = 20$			
		Fit	Sim	Bias	RMSE	Fit	Sim	Bias	RMSE
a	Mean	1.121	1.193	-0.072	0.189	1.125	1.243	-0.119	0.069
	SD	0.477	0.781	0.384	0.336	0.463	1.142	0.877	0.053
b	Mean	0.000	0.000	0.000	0.138	0.000	0.000	0.000	0.154
	SD	0.827	0.841	0.302	0.138	0.908	0.963	0.341	0.257
c	Mean	0.267	0.262	0.006	0.005	0.272	0.234	0.038	0.008
	SD	0.060	0.060	0.063	0.008	0.070	0.070	0.077	0.009
θ	Mean	0.009	0.018	-0.009	0.388	0.000	-0.033	0.033	0.406
	SD	0.593	0.995	0.462	0.431	0.588	0.999	0.468	0.352
η	Mean	0.506	0.669	-0.163	0.061	0.505	0.678	-0.173	0.064
	SD	0.085	0.175	0.163	0.056	0.084	0.175	0.16	0.063
λ	Mean	1.469	1.411	0.059	2.027	1.472	1.397	0.076	1.606
	SD	0.246	1.451	1.391	8.715	0.241	1.258	1.226	5.205

		$n = 500,\ \ k = 40$				$n = 1000,\ \ k = 40$			
		Fit	Sim	Bias	RMSE	Fit	Sim	Bias	RMSE
a	Mean	1.064	1.149	-0.085	0.123	1.055	1.149	-0.094	0.095
	SD	0.392	0.737	0.477	0.197	0.414	0.737	0.421	0.137
b	Mean	0.000	0.000	0.000	0.112	0.000	0.000	0.000	0.499
	SD	1.019	1.069	0.238	0.104	2.093	1.069	1.473	0.891
c	Mean	0.279	0.267	0.012	0.006	0.297	0.267	0.030	0.008
	SD	0.070	0.059	0.069	0.006	0.081	0.059	0.077	0.007
θ	Mean	0.009	-0.016	0.025	0.337	0.003	-0.015	0.019	0.329
	SD	0.657	0.990	0.402	0.353	0.676	1.001	0.401	0.365
η	Mean	0.521	0.667	-0.146	0.057	0.516	0.668	-0.152	0.056
	SD	0.122	0.181	0.153	0.062	0.122	0.174	0.148	0.055
λ	Mean	1.408	1.316	0.092	1.338	1.435	1.356	0.078	1.328
	SD	0.357	1.161	1.094	5.997	0.376	1.180	1.090	3.229

Table 2 - Model comparison for the NWF data by considering a *3PL* and a *alternative 3PL speeded* IRT model

Model	Number of parameters	Dbar	DIC
3PL	662	9083	9459
alternative 3PL speeded	1686	8314	8886

Table 3 - Descriptive statistics and correlations between personal latent variables under the *3PL*, the *alternative 3PL speeded* IRT model and raw scores for 512 Peruvian students in the NWF test

Models	Personal variables	Descriptive statistics				correlations				
		Mean	Min.	Max.	S.D.	θ	θ^*	η	λ	raw scores
3PL	θ	-0.047	-3.745	2.405	1.473	1.000				
alternative	θ^*	-0.016	-2.506	2.293	1.239	0.725	1.000			
3PL speeded	η	0.568	0.007	0.862	0.243	0.874	0.446	1.000		
IRT	λ	3.328	0.229	27.550	5.116	-0.727	-0.136	-0.712	1.000	
Classical	raw score	30.701	0.000	50.000	12.976	0.996	0.733	0.879	-0.707	1.000

θ : ability under the *3PL*, θ^*: ability under speededness, η: tolerance toward speededness, λ: propensity to guessing under speededness and raw scores: number of words correctly read.

Table 4 - Personal latent variables for some low raw scoring Peruvian students in the NWF test under the alternative 3PL speeded model

case	θ^*	η	λ	raw scores	pattern of response
234	-0.201	0.187	7.630	16	11111111110111111000000000000000000000000000000000
185	-0.214	0.027	12.360	4	10110100000000000000000000000000000000000000000000
276	-0.196	0.010	20.130	1	01000000000000000000000000000000000000000000000000
195	-0.208	0.008	20.790	1	00010000000000000000000000000000000000000000000000
257	-0.192	0.008	20.930	1	00010000000000000000000000000000000000000000000000
172	-0.198	0.007	26.890	0	00000000000000000000000000000000000000000000000000

The group, in general, has a low tolerance towards speededness and, as expected, a high propensity to guessing. Although many of the raw scores and measures of abilities under the *3PL* were different, we find similar measures under the alternative *3PL speeded* model. Note that, this is only true for low scores. Consequently the *alternative 3PL speeded* model assigns similar NWF measures to low scoring examinees and is able to differentiate among examinees in terms of their tolerance and propensity due to a speededness situation.

Finally, in Figure 6 we show the discrimination, difficulty and guessing estimates under the *3PL* and the *alternative 3PL speeded* IRT models. In general these estimates look very similar, but there are remarkable differences in some items due to the speededness phenomenon. For instance, the estimates of the difficulty parameters under the *3PL* are larger than under the *alternative 3PL speeded* model. This confirms the statement of Oshima (1994) presented in the introduction of section 3 that under speed conditions last items tend to appear harder than they are when administered under non speededness conditions. Moreover, we found that several items, especially the first, present high values for the discrimination parameter under a *3PL*. Therefore, speededness seems to overestimate the discrimination power of these items.

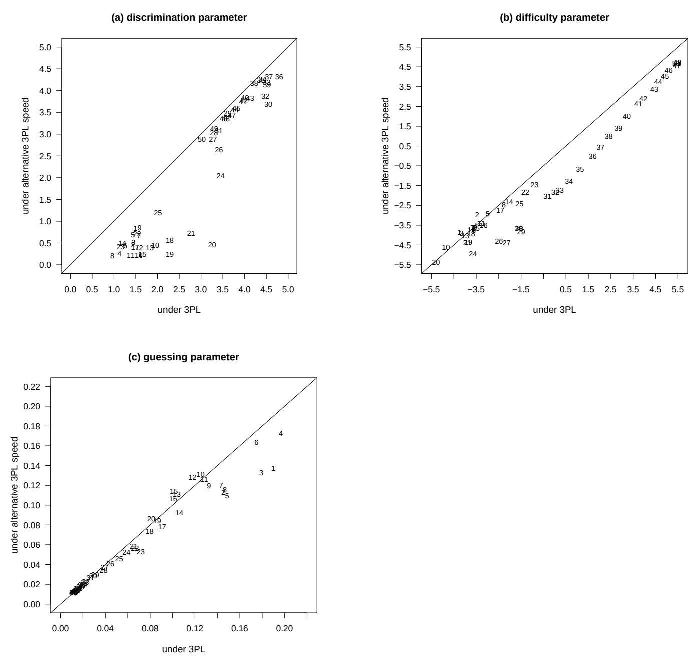

Figure 6 - Item parameter estimates under the *3PL* and the *alternative 3PL speeded* IRT model.

6 Conclusions

The speeded item response model with gradual process change of Goegebeur *et al.* (2008) is an interesting model to analyze data with evidence of speededness offering a new set of personal latent variables to better understanding the performance of the examinees in a test. Inspired in the classical approach of this model, we propose an *alternative 3PL speeded* IRT model which initially consider an independent structure for the personal parameters under Bayesian approach. Independent priors are assumed not only for the personal parameters, but also for the item parameters as is usually considered in the Bayesian framework. Following this specification, dependence between the parameters in the model can be obtained in the posterior distribution.

In Bayesian IRT models the priors have an additional role to identify the model and not only to give an idea about the uncertainty of the parameters in the model. For example, when we consider b and θ to be normally distributed with mean 0 and variance 1, we are also giving a simple scale to these parameters which is easy to interpret. By considering the results of our simulation study, we suggest to perform a further prior sensitivity analysis for the propensity to guessing λ parameter.

The model can be easily implemented using a hierarchical formulation in several packages as WinBUGS, OpenBugs, SAS, Rstan or JAGS. Although there is a large number of parameters in the model, our estimators using WinBUGS performed reasonably good and they were obtained in a straightforward way.

We only observe some bias in the estimation of the λ parameter and consequently a further research is needed to study the sensibility of this parameter to alternative prior specifications. More comprehensive studies are needed to obtain conclusive results. Additionally, we note that the term $P_j(\eta_i, \lambda_i)$ in the probability of a correct response in (4) is a penalizing factor under speededness and consequently alternative models can be formulated considering other penalizing factors with different rationality.

As a future work, we suggest to explore others models for the EGRA data, such as, the one given by Jansen (1997) or the one proposed by van der Linden (2006) for speeded tests with response times. To go in the last direction we would additionally requiere the subject response times for the items. In addition a further sensitivity study for the priors and other Item Characteristic Curves, such as the skew ICC given by Bazan *et al.*, (2006), can be explored.

Acknowledgments

The first author was partially supported by FAPESP (2017/07773-6). Luis Valdivieso was supported by PUCP- DGI - 2016-1-0077/ ID 342 and DAI - 2009 – 033. The authors thank the Editor, the Associate Editor, and the Referees for useful comments, which improved the original version of the article.

BAZÁN, J. L.; VALDIVIESO, L.; BRANCO, M. D. Measurement of the nonsense word fluency: A Bayesian approach to a item response model with speededness. *Rev. Bras. Biom.,* Lavras, v.35, n.4, p.810-833, 2017.

- *RESUMO: Aceleração refere-se à situação em que o limite de tempo num teste padronizado não permite a um número substancial de examinandos poder fazer todos os itens do teste, assim, estimativas usando modelos de resposta ao item logísticos de tres parâmetros (3PL) conduz a estimativas contaminadas dos parâmetros do modelo. Este trabalho propõe um modelo bayesiano simples para estimar ambos os tipos de parâmetros: parâmetros pessoais e parâmetros de item para dados de um teste com evidência de aceleração. O modelo está fortemente relacionado com um modelo proposto por Goegebeur, De Boeck, Wollack e Cohen (2008) mas diferente deste, a estrutura de dependência nos parâmetros pessoais não é assumida inicialmente. Nós conduzimos um estudo de caso para analisar dados de fluência em palavras sem sentido em estudandes peruanos, o qual presenta evidencia de aceleração. Comparando os resultados destes dados com o $3PL$ e o modelo proposto, nós encontramos como esperado, que dificuldade e discriminação são superestimados sob $3PL$. Medidas semelhantes das habilidades dos examinandos em ambos modelos são obtidas e novos parâmetros pessoais: tolerância e propensão frente a aceleração são obtidos considerando-se o modelo proposto. Finalmente, futuros estudos são sugeridos.*

- *PALAVRAS-CHAVE: Modelos de resposta ao item; estimação bayesiana; aceleração; modelo logístico; variáveis latentes pessoais, fluência en palavras sem sentido.*

References

ALBERT, J. H. Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, v.17, p.251-269, 1992.

ALBERT, J. H. and GHOSH, M. Item response modeling. In Dey, K. D. , Ghosh, S. K.; Mallick B. K. (Ed.) *Generalized linear models: A Bayesian perspective.* New York: Marcel-Dekker, 2000. p.173-193.

AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, v.12, p.171-178, 1985.

BAKER, F. B.; KIM, S. -H. (2004). *Item response theory - Parameter estimation techniques.* 2ed. New York: Marcel Dekker, Inc., 2004. 528p.

BAZÁN, J. L.; BRANCO, M.; BOLFARINE, H. A skew item response model. *Bayesian Analysis*, v.1, p.861-892, 2006.

BAZÁN, J. L. Item response theory in WinBUGS using BayesianModeling software. Software highlight in *ISBA Bolletin*, v.19, n.1, p.8-10, 2012.

BÉGUIN, A. A.; GLAS, C. A. W. Estimation of multidimensional IRT models. *Psychometrika*, v.66, p.541-562, 2001.

BEJAR, I. H. *Test speededness under number-right scoring: An analysis of the Test of English as a Foreign Language*. (Report No. ETS-RR-85-11), 2006, 57p. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 263128).

BOCK, R. D.; AITKIN, M., Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm, *Psychometrika*. v.45, p.433-459, 1981.

BOLT, D. M.; COHEN, A. S.; WOLLACK, J. A. Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, v.39, p.331-348, 2002.

BOLFARINE, H.; BAZÁN, J. L. (2010) Bayesian estimation of the logistic positive exponent IRT model. *Journal of Educational Behavioral Statistics*, v.35-6, p.693-713, 2010.

CAI, L.; DU TOIT, S. H. C.; THISSEN, D. *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling* [Computer software]. Chicago: SSI International. 2011.

CURTIS, M. S. BUGS Code for Item Response Theory. *Journal of Statistical Software*. v.36, Snippet 1, p.1-34, 2010.

DE BOECK, P.; M. BAKKER, R.; ZWITSER, M.; NIVARD, A.; HOFMAN, F.; TUERLINC, K. X.; PARTCHEV, I. The estimation of item response models with the lmer function from the lme4 package in r. *Journal of Statistical Software*. v.39, n.1, p.1-28, 2011.

FIEN, H.; BAKER, S.; SMOLKOWSKI, K.; MERCIER, J.; KAME´ENEUI, E.; BECK, C. Using nonsense word fluency to predict reading proficiency in kindergarten through second grade for English learners and native English speakers. *School Psychology Review*. v.37, p.391-408. 2008.

FOX, J. P. *Bayesian item response modeling: Theory and applications*. New York: Springer, 2010. 314p.

FOX, J. P.; GLAS, C. A. W. Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, v.66, p.269-286, 2001.

FU, Z-H; TAO, J.; SHI, N-Z. Bayesian estimation in the multidimensional three-parameter logistic model. *Journal of Statistical Computation and Simulation*, v.79, n.6, p.819-835, 2009.

GELMAN, A.; HILL, J. *Data analysis using regression and multilevel/hierarchical Models*. Cambridge: Cambridge University Press, 2007. 625p.

GOEGEBEUR, Y.; DE BOECK, P.; WOLLACK, J. A.; COHEN, A. S. A speeded item response model with gradual process change, *Psychometrika*, v.73, p.65-87, 2008.

JACKSON, S. pscl: Classes and methods for R developed in the political science computational laboratory, Stanford University. Department of Political

Science, Stanford University, Stanford, California. R package version 0.95, http://CRAN.R-project.org/package=pscl.

JANSEN, M. The rasch model for speed test and some extensiones with applications to incomplete designs. *Journal of Educational and Behavioral Statistics*, v.22, p.125-140, 1997.

KIM, J. S.; BOLT, D. M. Markov chain Monte Carlo estimation of item response models. *Educational Measurement: Issues and Practice*, v. 26, p. 38-51. 2007.

KLEIN ENTINK, R. H.; FOX, J. P.; VAN DER LINDEN, W. J. A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, v.74, p.21-48, 2009.

LU, Y.; SIRECI, S. G. Validity issues in test speededness. *Educational Measurement: Issues and Practice*, v.26, p.29-37, 2007.

MAIR, P.; R. HATZINGER. Extended rasch modeling: The erm package for the application of irt models in r. *Journal of Statistical Software*, v.20, n.1, p.1-20, 2007.

OSHIMA, T. C. The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement*, v.31, n.3, p.200-219, 1984.

PATZ, R. J.; JUNKER, B. W. A straightforward approach to Markov Chain Monte Carlo methods for item response models. *Journal of Educactional and Behavioral Statistics*, v.24, p.146-178, 1999.

RIZOPOULOS, D. ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, v.17, n.5, p.1-25, 2006.

RUPP, A.; DEY, D. K.; ZUMBO, B. To Bayes or Not to Bayes, from Whether to When: Applications of Bayesian methodology to item response modeling. *Structural Equations Modeling*, v.11, p.424-45, 2004.

RTI-FDA. *Snapshot of School Management Effectiveness.Peru Pilot Study*. 2008, 53p,. EDDATA II. RTI. Available in http://pdf.usaid.gov/pdf_docs/Pnadp642.pdf.

SAHU, S. K. Bayesian estimation and model choice in item response models. *Journal of Statistical Computation and Simulation*, v.72, p.217-232, 2002.

SINHARAY, S.; JOHNSON, M. S.; STERN, H. S. Posterior predictiv e assessment of item response theory models. *Applied Psychological Measurement*, v.30, p.298-321, 2006.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, Series B, v.64, n.4, p.583-616, 2002.

STAFFORD, R. E. The speed quotient: A new descriptive statistic for tests. *Journal of Educational Measurement*, v.8, p.275-278, 1971.

STONE, C. A.; ZHU, X. *Bayesian analysis of item response theory models using SAS*. SAS Institute Inc. Cary, NC, USA. 2015, 265p.

VAN DER LINDEN, W. J. A lognormal model for response times on test items. *Journal of Educactional and Behavioral Statistics*, v.31, p.181-204, 2006.

WANG, X., BERGER, J. O.; BURDICK, D. S. Bayesian analysis of dynamic item response models in educational testing. *The Annals of Applied Statistics*, v.7, p.126-153, 2013.

WOLLACK, J. A.; COHEN, A. S. *A model for simulating speeded test data*. Technical Report. 2005, 29p. University of Wisconsin–Madison, Madison.

YAMAMOTO, K.; EVERSON, H. Modeling the effects of test length and test time on parameter estimation using the hybrid model. In Rost, J. and Langeheine, R. (Eds.). *Applications of latent trait and latent class models in the social sciences*. New York: Waxmann, 1997, p.89-99.

Received on 23.03.2016.

Approved after revised on 30.12.2016.

Appendix

```
#Alternative 3PL speed model by considering
#bayesian parameterization in WinBUGS

model {
  for (j in 1 : k) {
  for (i in 1 : n) {
# latent structure
      m[i,j]<- a[j]*theta[i] -b[j]

# ideal probability without speed and guessing
    p2l[i, j]<-exp(m[i, j])/(1+exp(m[i, j]))
# decay
       d[i,j]<- (j/k)-eta[i]
    deca[i,j]<- pow(1-d[i,j],lambda[i])
# speed probability (penalization)
 pspeed[i,j]<- min(1,deca[i,j])
# ideal probability with penalized factor
  p[i, j]<-c[j]+(1-c[j])*p2l[i, j]*pspeed[i, j]
   y[i, j] ~ dbern(p[i, j])
                  }
                  }
#independent priors for persons parameter
   for (i in 1:n) {

 theta[i] ~ dnorm(0,1)
 # eta[i]~ dbeta(2,2)
   eta[i]~ dbeta(1,1)#Uniforme
 lambda[i]~ dlnorm(0,1)
                 }
#priors for items parameter
   for (j in 1:k) {

#Bazan et al (2006)
    a[j] ~ dnorm(1,2)I(0,)
    b[j] ~ dnorm(0,0.5)
#Fu, Tao and Shi (2009)
    c[j] ~ dbeta(5,17)

  bv[j]<- b[j]/a[j]
  bc[j]<- bv[j] - mean(bv[])
                 }
}
```

```
data
list(n=512,  k=50)
#load your data in other file

Inits
list(a=c(1,..,1),
b=c(0,...,0),
c=c(0.2,...,0.2),
theta=c(
0.5,...,0.5),
eta=c(0.5,...,0.5),
lambda=c(
0.5,...,0.5,))
```