BAYESIAN ANALYSIS IN A SPEEDED ITEM RESPONSE MODEL

J. BAZAN, L. VALDIVIESO*, M. D. BRANCO [1]**

∗ Departamento de Ciencias, Pontificia Universidad Católica de Lima, Perú
∗∗ Departamento de Estatística, Universidade de São Paulo, São Paulo, Brazil

April 30, 2010

Abstract

We develop a Bayesian analysis of the model proposed by Goegebeur, de Boeck, Wollack and Cohen (2008) to work with speededness by considering a gradual process change IRT model. Prior specification and MCMC algorithm are discussed. Simulations studies are conducted to evaluate his performance to the recover the parameters by considering independence or dependence in examinee´s parameter or personal latent variables by using independent priors. A real-data example, EGRA test, is analyzed and show how the model can be used to identified personal latent variables associated with response in Speededness situations which is consistent with the results in the test.

Keywords: *Item response models, Bayesian estimation, Speededness, guessing parameter, logistic model, DIC, simulation, EGRA* .

[1] *Address for correspondence:* Jorge Luis Bazan. Av Universitaria 1801, San Miguel, Lima 32, PUCP, Peru. E-mail: jlbazan@pucp.edu.pe.

1 Introduction

Test speededness may be defined as the degree of which examinee do not have sufficient time to respond to all test items (questions). As indicated by Lu and Sireci (2007), speededness refers to the situation where the time limit on a standardized test do not allow substantial number of examinees to fully consider all test items. Bejar (1985, p.1) says "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".

Speededness in a Test introduces a severe threat to the validity of interpretation based on test scores when tests are not intended to measure speed of responding. In fact, as indicated by Lu and Sireci (2007), an underlying assumption in a test is that scores accurately reflect the trait the test is intended to measure and the differences in the scores obtained by various examinees represent differences in the degree to which they possess or lack that trait. When speededness in unintended, this introduces construct-irrelevant variance into the test scores and thus change the construct the test intends to measure. Consequently, the scores obtained from the test no longer provide an adequate basis for the kinds of inferences the test user wishes to be able to make.

In the context of Unidimensional Item Response Theory (IRT) the probability that an examinee gives a correct response to an item depends only on the examinees´s proficiency and the characteristics of the item. In addition there is a IRT assumption of that the examinee have sufficient time to answer all items in the 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 assumption of sufficient time is violated as 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).

A clear example of this phenomena can be illustrated in Figure 1 where the percentage of students able to read correctly nonsense words decreasing the end of the test then items in the end can be considered difficulty. The Figure 1 was obtained from a pilot study to adapt the Snapshot of School Management Effectiveness (SSME) and Early Grade Reading Assessment (EGRA) instruments to the Peruvian reality (FDA, 2008). The study was supported by the Research Triangle Institute (RTI) to the Foundation for Agricultural Development. The EGRA instrument includes a battery of test that are applicable for the initial years of education in primary school and provides a quick information regarding school management and pre-reading skills. Nonsense word test, described in section 5.1, is a test where the students have to read 50 nonsense words with a time limit of 60 seconds. In this case Speededness can be easily detected as the percentage of correctly read words tends to decrease conforms the location of the words approaches to the end of the reading sequence.

As indicated by Yamamoto and Everson (1997), traditional methods of assessing test speededness are limited to analysis of distributions of missing responses, especially at the end of a test. However, analysis of missing responses is inadequate in evaluating the speededness of a multiple choice test when the test

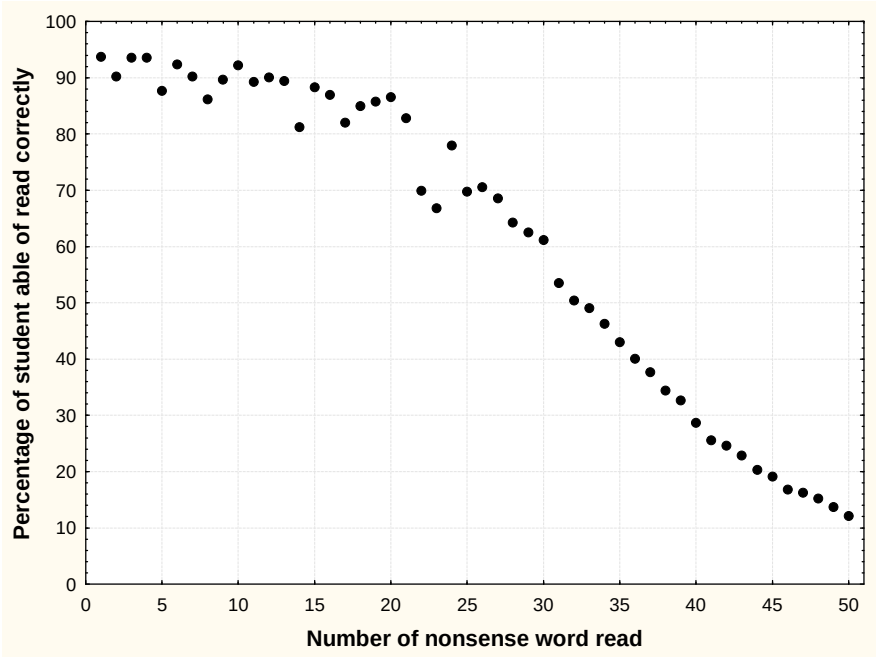

Figure 1: Percentage of correctly read nonsense words.

score is a function of the total number of correct response. A revision of this methods can be seen in Lu and Sireci (2007).

In the past few years, several models have been developed that account for local dependencies due to speededness effects. Yamamoto and Everson (1997) have proposed a hybrid model which assumes that a two parameter logistic IRT model, named here as 2L, is appropriate throughout the most of the item is the test. However the last items are answered randomly by some subset of examinees. Bolt, Cohen and Wollack (2002) have proposed a mixture Rash model which assumes that two latent classes of examinees exist, and the Rash model is appropriate for both classes. For items early in the test, the model regards the item difficulty parameters to be equal in both classes; however for the last items, the corresponding parameters are constrained to be larger that the initial one. In this paper we adopt a more flexible speeded model, recently developed by Goegebeur, *et al* (2008) called speeded item response model with gradual process change and named here 3Lspeeded since this is a three item parameters logistic IRT model. Also it assumes three dependent personal latent variables to explain the behavior of the examinees in the test under the speeded situation. Thus the 3Lspeeded is a multidimensional IRT model. Take in account another examinees´s parameters or personal latent variable is in concordance with the indicated by Yamamoto and Everson (1997, p.1), "the speed of performing a task is one of the more noticeable aspects in which individuals differ from each other, in addition to the differences in the abilities to perform a task correctly".

In this paper we adopt a total Bayesian perspective using MCMC methods in contrast to the marginal maximum likelihood approach proposed by the authors in the estimation of the 3Lspeeded IRT model.

Moreover, an short discussion about of prior specification is conduced. We apologize that independent priors are sufficient to take in account the dependence structure of examinees´s parameters and thus specification of the dependence by considering copulas in the original version is unnecessary. In addition, we conduce several simulations studies about of the power of the approach to recovery of parameters, misspecification of the model and evaluation of the dependence by considering independent priors. Finally we development an application to the EGRA test data shown in Figure 1.

The paper is organized as follows. In section 2 we make a review of the 3Lspeed IRT model with emphasis in examinees´s parameter or personal latent variables interpretation. In the third section, the Bayesian estimation approach is developed using the WinBUGS software with a discussion about the prior choices and their independence. In section 4 fourth simulation studies are performed. These are designed to, respectively, recover the parameters in the model by considering independence and dependence personal latent variables and to evaluate the misspecification of the model. Section 5 presents an application to the EGRA data showing the importance of considering speededness effects in comparison with usual IRT models. We conclude in Section 6 by suggesting some extensions to the 3Lspeed model and future works.

2 Model

IRT models to dichotomous item responses assumes that the sequence of binary random variables $\{Y_{ij} : 1 \leq i \leq n\,;\, 1 \leq j \leq k\}$ associated with item responses are conditionally independent given θ_i , the latent variable associated with the ability or latent trait for individual i. Here $Y_{ij} = 1$ if examinee i correctly answers item j and $Y_{ij} = 0$ otherwise. The response pattern associated with the examinee i is written as $Y_i = (Y_{i1}, \ldots, Y_{ik})$. The IRT model assumes that the probability of a correct response is given by

$$p_{ij} = P[Y_{ij} = 1 \,|\, \theta_i \,,\, a_j \,,\, b_j] = F(m_{ij}), \tag{2.1}$$

F is called the item characteristic curve (ICC) and

$$m_{ij} = a_j(\theta_i - b_j),\ i = 1, \ldots, n,\ j = 1, \ldots, k \tag{2.2}$$

is the latent linear predictor involving the discrimination item parameters a_j, the difficulty item parameters b_j and the values of the latent variable θ_i associated to the examinees´s i ability. High values of ability θ mean that a person have a major probability of correct response.

Two popular examples for the ICC are the cumulative distribution function (cdf) of the standard normal distribution and the standard logistic distribution. Such models are called the 2P and 2L IRT models by emphasizing the use of normal and logistic cdf, respectively. By adding a guessing item parameter c_j, one obtains the 3L IRT model

$$p_{ij} = P[Y_{ij} = 1 \,|\, \theta_i \,,\, a_j \,,\, b_j,\ \ c_j] = c_j + (1 - c_j)F(m_{ij}). \tag{2.3}$$

Recently Goegebeur *et al* (2008) have formulated a speeded item response model with gradual process change named here as 3lspeeded IRT model. This model assumes that the probability of a correct response is given by

$$p_{ij} = P[Y_{ij} = 1 \,|\, \theta_i \,,\, \eta_i \,, \lambda_i \,,\, a_j \,,\, b_j,\ \ c_j] = c_j + (1 - c_j)G(m_{ij}), \tag{2.4}$$

where

$$G(m_{ij}) = F(m_{ij})P_j(\eta_i, \lambda_i),$$
$$P_j(\eta_i, \lambda_i) = min\Big\{1, r_j(\eta_i, \lambda_i)\Big\}, \tag{2.5}$$

and

$$r_j(\eta_i, \lambda_i) = \Big[1 - (\frac{j}{k} - \eta_i)\Big]^{\lambda_i}. \tag{2.6}$$

Note that, two additional parameters are introduced in this model, for each examinee, $\eta_i \in [0, 1]$ and $\lambda_i > 0$, where η_i can be will be called the *Tolerance towards Speededness* and λ_i can be will named *Propensity to guessing in Speededness situation.* The parameter η_i is introduced in the model to identify the point, j/k, expressed as a fraction of the number of items or relative position of the item, where examinee i first experiences an effect due to speededness. The parameter λ_i control the rate of decrease towards a guessing situation.

The rationality behind this model is as follows. When examinee i encounters item j, the examinee answers according to a 3L IRT model or a random guessing process, with probabilities $P_j(\eta_i, \lambda_i)$ or $1 - P_j(\eta_i, \lambda_i)$ respectively. Under the problem solving process the examinee knows the answer with probability (2.1); if ignorant, the examinee guesses at random.

$P_j(\eta_i, \lambda_i)$ in (2.5) can be consider as a *penalize factor in speededness situation.* If there is not speededness situation, that is if a person i have a value of tolerance toward speededness η_i major o equal that the relative position of item j/k, the distance between both in (2.6) is negative and then the value of $r_j(\eta_i, \lambda_i) > 1$ and thus $P_j(\eta_i, \lambda_i) = 1$ and consequently $F(m_{ij})$ is no penalized. In another case, if a person i have a value of tolerance η_i minor that that the relative position of item j/k, then exist a speededness situation, the distance between both in (2.6) is positive and then the value of $r_j(\eta_i, \lambda_i) < 1$ and thus $P_j(\eta_i, \lambda_i) < 1$. In this case $P_j(\eta_i, \lambda_i)$ penalize (decrease) the probability of correct response of the item to this person $F(m_{ij})$. In addition this factor of decrease can be accelerated by considering high values of the examinee´s parameter λ_i.

The parameter η_i can be consider as a personal latent variable associated with the tolerance to speededness in contrast with the relative position the item in the test j/k. This conduct is quite similar with the conduct of θ (ability) and b (difficulty), one associated with the person and another associated with the item. η_i and j/k, both should be in the same scale (between 0 and 1). However η_i is a random variable and j/k is fix. High values of η_i mean that the person cannot be in speededness situation and then your probability $F(m_{ij})$ is not penalize by $P_j(\eta_i, \lambda_i)$. When η_i is near of 1 then we expected that the speededness point parameter indicate that there is a situation of speededness to end of the test. When η_i is near of zero, then the speededness point mean indicate that there is a situation of speededness to begin of the test.

High values of λ_i, when there is a speededness situation, mean that the penalization $P_j(\eta_i, \lambda_i)$ of the probability of correct response is incremented (decrease slowly the probability of correct response p_{ij}). Note that, if λ_i is high but there is not speededness situation because the tolerance is adequate, then the penalization not occur. Thus, η_i parameter is more important that λ_i parameter because your interpretation exist in any scenario. In the other hand the interpretation for λ_i is only possible

under speededness situation. Thus λ_i parameter express a propensity or tendency to guessing which no necessarily affect the probability of correct response.

3 Bayesian Estimation

The likelihood function for model (2.4) is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\Theta}|\boldsymbol{y}) = \prod_{i=1}^{n} \prod_{j=1}^{k} p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}}, \tag{3.1}$$

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 examinee´s 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 of the n examinees to the k items.

The approach used in Goegebeur *et al* (2008) to the estimation of the model is marginal maximum likelihood by considering examinees´s parameters as random effects and integrating out and the resulting likelihood is maximized with respect to item parameters. This algorithm is implemented in NLMIXED procedure in SAS by considering gaussian quadrature for approximating the integrals needed. In addition, the estimation of examinee´s parameters is performed in a second stage with item parameters replaced by estimates computed previously including empirical Bayesian estimation, Bayesian marginal estimation using the posterior mode or mean for examinee´s parameters. Limitations of this methodology for IRT models are discussed in Patz and Junker (1999) and Sahu (2002).

In contrast we adopt in this paper a Full Bayesian estimation methodology by considering MCMC estimation. Several researchers (see, for example, Swaminathan et al., 2003) have demonstrated that accurate item estimation parameters in small samples can only be accomplished through a Bayesian approach.

3.1 Priors specification

For simplicity, we assume independent priors for the whole set of parameters. Empirical evidence (see Patz and Junker, 1999, among others) suggests the presence of posterior correlation between item parameters. To take this into account, we could consider a dependence structure for the priors. This is not only difficult to implement, but also involves the increment in the number of parameters for the model. Hence, we prefer to use independent and common priors for $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ and let such correlations be only data dependent. Goegebeur, *et al*(2008) have proposed the use of a normal copula for modeling the dependence between examinee´s parameters. We consider that this level of complexity turns out to be unnecessary in the Bayesian approach, as we will see in section 4.1

The following usual priors in IRT model are specified for item parameters :

$$a_j \sim Lognormal\,(0, 0.5^2) \; or \; Normal(0, 0.5)I(a > 0),\; b_j \sim Normal\,(0, 1),\;\; c_j \sim Beta\,(5, 17).$$

$Lognormal(0, 0.25)$ (Patz and Junker, 1999) and $N(0, 0.5)I(a > 0)$ (Bazan, et al, 2006) has $\mu = 1.13$ and $\sigma = 0.60$ and $\mu = 1.11$ and $\sigma = 0.61$ respectively and a sampled value from these distributions

are always bigger than 0. Likewise, $Beta(5,7)$ (Fu, Tiao and Shi, 2009) has $\mu = 0.23$ and $\sigma = 0.09$. $Normal(0,1)$ for b distributions are give by example Albert and Ghosh (2000) or Sinharay and Johnson (2003).

Give a person i, each person have three latent personal that explain our response in the test under the present model. One is ability (θ_i), another is tolerance toward speededness situation (η_i) and the third is the impact of the speededness situation in your tendency to guessing or propensity to guessing in speeded situation (λ_i). The values of θ are give in $\Re$ but is expected values between $[-3,3]$ to the most of the times. Values of η are given in [0,1] and values of λ are give in $\Re$.

Following Goegebeur, *et al*(2008) we will set the following priors for personal latent variables :

$$\theta_i \sim N\,(0,1),\ \eta_i \sim Beta\,(\alpha,\beta),\ \ \lambda_i \sim LoNormal\,(mu_\lambda, \sigma^2_\lambda).$$

It is actually common in IRT modeling assume $Normal\,(0,1)$ as prior, for θ_i, $i = 1,\ldots,n$ which contribute in the identification of the model. Also it natural to consider a beta prior to η_i parameter, that is $\eta_i \sim Beta\,(\alpha,\beta)$. Specification to α and β can be choose by considering information about of the point more probably for speededness obtained of judges of experts about of the mean an variance of η_i. For instant a choose can be $Beta(4,2)$ that has $\mu = 0.67$ and $\sigma = 0.18$, which can be considered when speededness point is moderated. Others possibilities are $Beta(2,9)$ or $Beta(9,2)$ that has $\mu = 0.18$, $\sigma = 0.01$ and $\mu = 0.82$, $\sigma = 0.01$ respectively. In the first case speededness point is expected severe and in the second case is expected mild. In situations where any information about of this speededness point is not available a $Beta(1,1)$ correspondent to a $U(0,1)$ could be used.

We call attention that for the application of Goegebeur *et al* (2008) by considering the estimated values of $\mu_\lambda = -3.604$ and $\sigma_\lambda = 2.771$ are reported values of λ for examinees with big variability as 3.417×10^{-6} and 22.4043. In addition by simulating 1000 samples generate of this distribution we found a maximum value of 17140.48 and minim value of 2.497×10^{-8}. We consider that the scale of values for this personal latent variables is bigger and the distance between values is not interpretable. An election possible to obtain a minor scale is consider $\mu_\lambda = 0$ and varying the value of σ_λ. By example if $\sigma_\lambda = 1$ is considered we obtain $\mu = 1.65$ and $\sigma = 2.16$. Also another election when a major variability can be considered is take $\sigma_\lambda = 0.8$ that has $\mu = 1.38$ and $\sigma = 1.3$ or $\sigma^2_\lambda = 0.8$ with $\mu = 1.49$ and $\sigma = 1.65$.

A Bayesian inference can be easily conducted by implementing a MCMC algorithm in the OpenBUGS environment.

4 Simulation studies

4.1 Recovery of parameters

To assess the performance of the proposed Bayesian estimation methodology we simulate a data set from the speeded model (2.4) by considering $n = 500$ examinees and $k = 20$ items under a logistic ICC. We consider the usual parametrization in Bayesian Inference (see by example Sahu (2002)) $m_{ij} = a_j\theta_i - bs_j$ where $bs_j = a_j \times b_j$ is the negative of a intercept parameter.

In order to simulate the data set we consider item parameters $\boldsymbol{a}$, $\boldsymbol{bs}$ and $\boldsymbol{c}$ generated from a $LogNormal(0, 0.5^2)$, $Normal(0, 1)$ and $Beta(12.5, 37.5)$ distributions respectively. For examinees´s parameters $\boldsymbol{\theta}, \boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ we consider two situations, one assuming this parameters independent and another case assuming that there is one dependence structure between they.

a) Independence between examinees´s parameters: in this case examinee parameter, $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ were generated from a $Normal(0, 1)$, $Beta(4, 2)$ and $LogNormal(0, 0.8^2)$ distributions respectively assuming independence between they.

b) Dependence between examinees´s parameters: For the second case, we consider that structure of correlations give in the application of Goegebeur *et al* (2008), that is, we consider that $\rho_{\theta\eta} = 0.659$, $\rho_{\theta\lambda} = -0.178$ and $\rho_{\eta\lambda} = 0.3199$ with variances equals to 1. In addition by considering a normal copula, a multivariate normal of dimension 3 was generated considering a mean vector of (0,0,0) and the matrix of correlations indicated above. Then "true" examinees´s parameters $\boldsymbol{\theta}, \boldsymbol{\eta}$ and $\boldsymbol{\lambda}$ were recovery by considering inverse distribution values under a $Normal(0, 1)$, $Beta(4, 2)$ and $LogNormal(0, 1)$ distributions, respectively.
The mean and variance of the simulated values in both cases are showed in Table 1 (independent case) and Table 2 (dependence case) respectively.

In both case, we consider the following structure of priors $a_j \sim N(1, 0.5)I(0,)$, $bs_j \sim Normal(0, 2)$, $c_j \sim Beta(5, 17)$, $\theta_i \sim Normal(0, 1)$, $\eta_i \sim Beta(1, 1)$, $\lambda_i \sim LogNormal(0, 0.8)$, where $b_i = bs_i/a_i$ is the difficulty parameter and bs_i is a intercept parameter used to compare we results with the estimates for the item difficulties in Goegebeur *et al* (2008). We considered a chain size of 1000 (20000 iterations with Bur-in=4000 and thin=20) in OpenBUGS. The results are shown in Table 1 and 2.

Table 1: Parameter recovery for the speeded IRT model by considering independence between examinees´s parameter

Parameter	simulated		fitted		MSE	RMSE
	mean	sd	mean	sd		
a	1.1345	0.5441	1.1499	0.4111	0.1599	0.3999
b	0.0925	0.7268	-0.0310	1.0376	0.2024	0.4498
c	0.2489	0.0744	0.2546	0.0574	0.0038	0.06158
θ	-0.02464	1.0148	-0.00029	0.7307	0.4354	0.6599
η	0.6484	0.1936	0.5038	0.1303	0.0642	0.2533
λ	1.3486	1.2156	1.4693	0.4291	1.5447	1.2428

In table 1, we found good estimations for a, b, η and c, acceptable ones for θ and some bias in the variability for λ which is underestimated.
By considering the results in Table 2 we found good estimations for all parameters with exception of λ parameter that present some bias in the variability which is underestimated.

Table 2: Parameter recovery for the speeded IRT model by considering dependence between examinees´s parameters

Parameter	simulated		fitted		MSE	RMSE
	mean	sd	mean	sd		
a	1.1345	0.5441	1.2125	0.4665	0.1597	0.3997
b	0.0925	0.7268	-0.1806	0.9225	0.2491	0.4991
c	0.2489	0.0744	0.2596	0.0711	0.0061	0.0783
θ	0.0026	1.0163	0.0031	0.7564	0.4054	0.6367
η	0.6612	0.1818	0.5092	0.1320	0.0525	0.2291
λ	1.6059	2.0871	1.4405	1.4405	4.4144	2.1010

Note here despite of that we generate in the second case of a $LogNormal(0,1)$ with $\mu = 1.65$ and $\sigma = 2.16$ different of $LogNormal(0, 0.8^2)$ with $\mu = 1.377$ and $\sigma = 1.30$ in the first case, we found bias λ parameter by considering a common prior give by $LogNormal(0, 0.8)$ with $\mu = 1.49$ and $\sigma = 1.65$. Subsequent studies by considering another priors or distributions for λ parameter may be directed to future works as discussed in final section.

Note also that in both simulations, there is a posterior dependence between the examinees´s parameters by each examinee since that by considering bayesian perspective we have a posterior conjoint distribution of this parameters by each examinee. The same is possible to item parameters as was observed by example Patz and Junker (1999). This makes it unnecessary to consider a copula for the dependence structure in the formulation of the model since the dependence of the parameters a posteriori is recovered by considering independent priors.

4.2 Misspecification of the Model

In order to detect the misspecification of unintended speed tests, we performed a comparative study. After generating two data sets from models (2.3) and (2.4) with a logistic ICC, we estimate their parameters. We opted for the Deviance Information Criterion (DIC) proposed by Spiegelhalter et al.(2002). This criterion does not penalize the number of parameters in the model and therefore is suitable to us, since the speeded model contains much more parameters than its embedded 3PL model.

Table 3: Mispecification of the speeded IRT model in comparison with the 3PL IRT model

Data Simulated	Criteria	Fit IRT Model	
		3PL (Non speeded)	Speeded
3L (Non Speed)	DIC	11740	11740
Speed	DIC	11880	11840

As showed in Table 3, the 3L and 3Lspeeded model fit the 3PL data. On the other hand, as expected,

the speeded data was better fitted by the 3Lspeeded model. A further research with more data sets is needed to confirm this pattern.

5 Application

5.1 EGRA example

An example of speedednees situation, indicated in the introduction, can be illustrated from a pilot study to adapt the Snapshot of School Management Effectiveness (SSME) and Early Grade Reading Assessment (EGRA) instruments to the Peruvian reality (FDA, 2008).

The EGRA instruments includes a battery of test that are applicable to educational institutes for the initial years of education in primary school and provide quick information regarding school management and "pre-reading" skills.

The next figure shows the items of one of the Phonics Test (Pseudoword Decoding or Nonsense Words) in which the students had to read 50 nonsense words with a time limit of 60 seconds.

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

In the Pseudoword Decoding subtest (Grades 1, 2 or 6 and 8 years old), the student must correctly identify nonsense words from a list, which assesses knowledge of letter-sound correspondences as well the ability to blend letters together to form unfamiliar nonsense.

Although Figure 1 suggests the presence of speededness, it is good to mention that non-reached words are considered here as incorrectly read. One way to measure speededness 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 naturally seems to be the case in our application, Stafford suggests to use the Speededness Quotation (SQ) index given by

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

where n is the number of examinees for the test, U_i denotes the number of items not reached by examinee i and W_i denotes the number of items incorrectly answered or omitted by examinee i. Indexes closer to 1 give evidence of speededness.

In the EGRA application we obtained a Speededness Quotation of 0.9081 which justify the use of the 3Lspeeded IRT model in comparison with usual 3L IRT model. For evaluate both model we consider the following structure of priors $\theta_i \sim Normal(0,1)$, $\eta_i \sim Beta(1,1)$, $\lambda_i \sim LogNormal(0,1)$, $a_j \sim Normal(1,0.5)I(0,)$, $bs_j \sim Normal(0,2)$, $c_j \sim beta(5,17)$ where $b_i = bs_i/a_i$ is the difficulty parameter and bs_i is a intercept parameter used to compare we results with the estimates for the item difficulties

in Goegebeur *et al* (2008).

Table 4 shows by considering DIC criteria, the Speeded model is better that 3L IRT model for the dataset of subtest of EGRA studied which confirm the evidence of SQ value.

Table 4: Model comparison to EGRA data by considering 3PL and speed IRT model

Simulated data	Number of parameters	$\bar{D}$	DIC	Computation time (hrs)
3L	662	9083	9459	2
3LSpeeded	1686	8314	8886	8

Table 5 shows the summary of the estimation of personal latent variables under 3lspeeded IRT model for the students in the sample in comparison with only latent variable under 3L IRT model. Also is include the score in classical approach.

Table 5: Summary of personal latent variables under 3Lspeeded and 3L IRT model including score

Model	Parameter	Mean	Minimum	Maximun	S.D.
3L	θ	-0.0471	-3.7450	2.4050	1.4731
3Lspeed	θ	-0.0162	-2.5060	2.2930	1.2387
	η	0.5679	0.0068	0.8621	0.2429
	λ	3.3279	0.2289	27.5500	5.1158
Classical	Score	30.7012	0.0000	50.0000	12.9755

In Figure 3 we show the correlation between different estimation to abilities by considering 3L and 3Lspeeded IRT model and classical score of the test. Note that the abilities under 3L model and classical score are high correlationated and then interchangeable. However it is not the case with the abilities under 3Lspeeded. In general the correlation with between abilities of 3L and 3Lspeeded IRT model are high but it is not true for all the examinees, especially to examinees with low values of ability where the Speededness situation can be more important since that the tolerance to Speededness situation can be low.

Table 6 shows the Pearson correlation coefficient among personal latent variables. Note that θ and η are positively correlationated which means that, when exist minor tolerance toward speeded then the ability tends to be minor but not for everyone. In addition, tolerance toward speeded and the Propensity to guessing in Speededness situation are negatively correlationated. This indicate that when major is the tolerance toward speeded then the personal effect in the speed situation is minor. Also, low negative correlation but significative, is observed between ability and personal effect in the speed

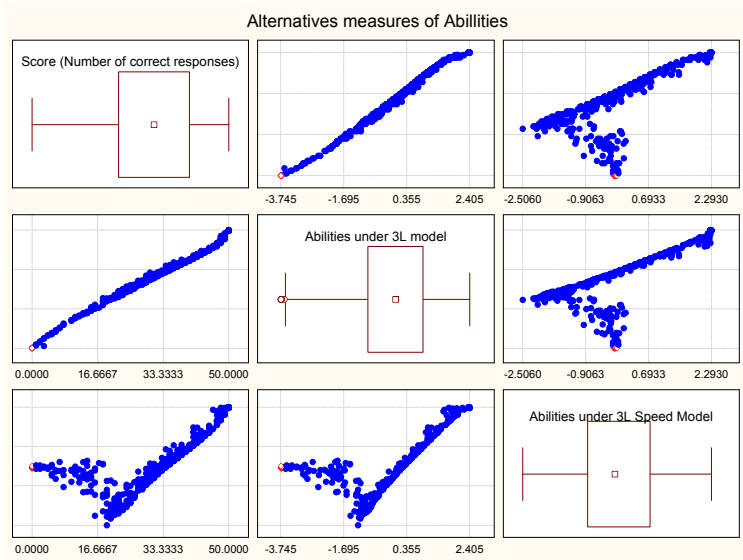

Figure 2: correlation in personal latent variables under speed IRT model including score

situation indicating that in some examinees when major is the propensity to guessing in Speededness situation then your ability tend to be minor. These results are showed in Figure 4. The importance of this type of results is that the examinees can be best characterized in relation to personal latent variables which is important in assessment.

Table 6: Correlations in personal latent variables under speed IRT model

	θ	η	λ
θ	1	0.446	-0.1365
η	0.446	1	-0.7117
λ	-0.1365	-0.7117	1

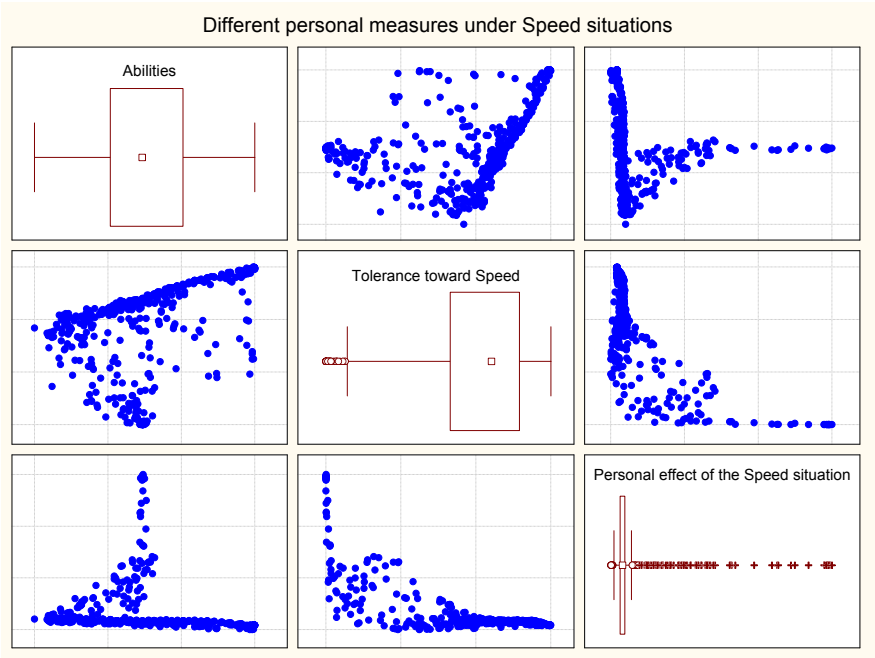

Figure 3: correlation in personal latent variables under speed IRT model including score

In figure 3, 4 and 5 we shows the discrimination, difficulty and guessing item parameters under 3L and 3lspeeded IRT model. In general the interpretation of the item parameters is the same under both models but in some items this is not true as is expected when exist a speed situation.

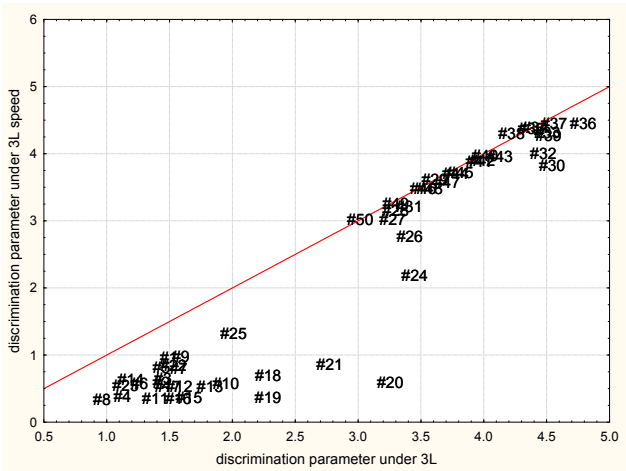

Figure 4: guessing item parameter under 3L and 3Lspeeded IRT models

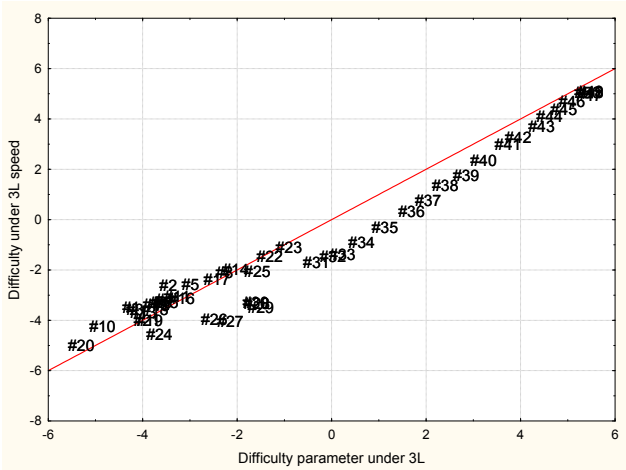

Figure 5: discrimination item parameter under 3L and 3Lspeeded IRT models

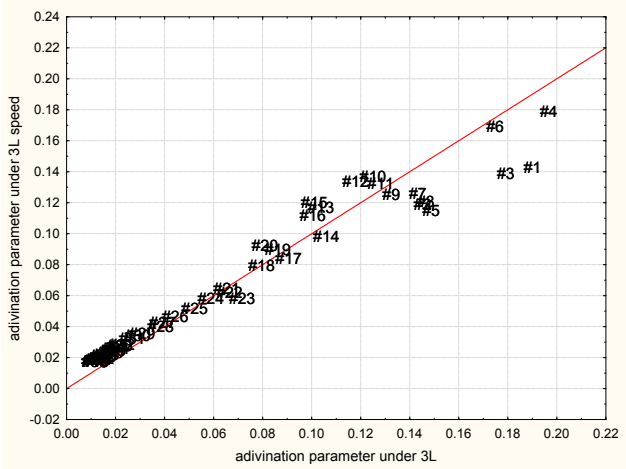

Figure 6: difficulty item parameter under 3L and 3Lspeeded IRT models

6 Final

In this paper we have presented the Bayesian estimation of the Goegebeur *et al* (2008) speeded item response model with gradual process change. This has been easily implemented using a hierarchical

formulation in OpenBUGS. Although the huge number of parameters in the model, our estimators performed reasonably good and are obtained in a considerable shorter time than using a marginal maximum likelihood approach. Although, the speeded model computation time is four times longer than the 3PL model, this time is much shorter than the one in the original report.

We have also shown that the speeded model is suitable not only when speededness is taken into account, but also when an embedded 3PL model without speededness generates the data.

A recovery parameter study has shown good results, but with a little bias in the λ parameter. We suggest an sensitivity study by considering some of the priors indicated here.

In IRT models the priors have an additional role to identify the model and not only give an idea about of our uncertainty under the parameters of the model. For example, when we consider β and θ with mean 0 and variance 1, we give a scale to the parameters in the model which is simples and can be interpretable. Following to Goegebeur, *et al*(2008) we have consider $\lambda_i \sim LogNormal\,(\mu_\lambda, \sigma^2_\lambda$ with $\mu_\lambda = 0$ and low values of σ^2_λ. We have found some bias and underestimation in the estimation of this parameter in simulations studies.

We call attention of the fact of that for λ parameter we only known that this is positive. Then several distributions with this restriction are possible. For example, exponential, gamma, weibull, half normal distributions also can be considered. In the case of consider gamma distribution extensive discussion is give about of this distribution as prior to positive parameters in Bayesian approach. By example a inverse-gamma(0.001,0.001) is usually a improper prior considered to the inverse of the parameter of interest. However as discussed by Gelman (2006) noninformative uniform priors in $[0, A]$ with A longer can be recommended to start.
However only is important that the distributions of λ conserve the order in the persons and can be easily interpreted. Thus, for ulterior studies we suggest prior specification of value of λ_i around of 1 to have a minor scale of values this parameter and facilitate your interpretation. Other simple prior that can be give to propensity to guessing in Speededness situation can be $\lambda_i \sim Exp(\beta_\lambda)$. We suggest for future works study the sensitivity of the model for different scenarios of priors.

In the application, extreme values of the lambda parameter are found with null responses. We will explore the possibility to delete such cases. As expected, the inclusion of speed in the model, has been found useful to find more reliable estimates for the ability and difficulty parameters.

In addition, by considering response times other models for speeded test as the given in Jansen (1997) and van der Linden (2006) can be explored for EGRA test.

References

[1] Bazan, J.; Branco, M.; Bolfarine, H. (2006). A Skew Item Response Model *Bayesian Analysis*, v.1,p. 861-892.

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

[3] Bolt, D.M.; Cohen, A.S.; Wolack, J.A. (2002), 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.

[4] FDA-USAID (2008). Snapshot of School Management Effectiveness.Peru Pilot Study. EDDATA II. RTI. Available in `https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail\&ID=152`

[5] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 1 515–533

[6] Goegebeur, Y.; De Boeck, P.; Wollack, J.A.; Cohen, A.S. (2008). A Speeded Item Response Model with Gradual Process Change, *Psychometrika*, v. 73, p. 65-87

[7] Jansen, M (1997). The Rascj Model for Speed Test and Some Extensiones With Applications to Incomplete Designs. *Journal of Educactional and Behavioral Statistics*, 22, p. 125-140

[8] Oshima, T. C. (1994). The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement*, 31(3), 200–219.

[9] Patz, R. J., Junker, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models. *Journal of Educactional and Behavioral Statistics*, 24, p. 146-178

[10] Spiegelhalter, D.J; N. G. Best; B. P. Carlin; A. van der Linde (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society*, Series B, 64.4 (2002) p. 583-616

[11] Swaminathan, H.; Hambleton, R; Sireci, S. ; Xing, D.; Rivazi, 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, p. 27-51

[12] van der Linden, W. J. (2006). A Lognormal Model for Response Times on Test Items. *Journal of Educactional and Behavioral Statistics*, 31, p. 181-204

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