MODERATING ABILITY OF ITEM RESPONSE THEORY THROUGH PRIOR GUESSING PARAMETER

Siow Hoo Leong, Siew Eng Ling and Rasidah Mahdi

Department of Mathematical Sciences, Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA, Kota Samarahan, Sarawak, Malaysia

ABSTRACT

A psycho-technology approach to discouraging guessing in multiple-choice formatted item can be done through reducing the a priori guessing probability of an item. This study proposes a psychometrics framework of Item Response Theory (IRT) to model the effect of having various priori guessing probabilities across different items. A prior guessing parameter is proposed to serves as a moderator of the ability parameter in the two parameter logistic IRT. The results show that the proposed prior guessing parameter successfully moderates the ability parameters of the subjects with different degrees of guessing. However, the prior guessing parameter is insensitive when the performance pattern is mixed within the testlet but similar across testlet with different priori guessing probabilities.

1. INTRODUCTION

The pioneering Item Response Theory (IRT) to deal with guessing in multiple-choice formatted item is the Parameter Logistic (3PL)-IRT model (Lord, 1980), where a guessing parameter is introduced to take account of the guessing effect on top of the difficulty and discrimination parameters in the 2PL-IRT model (Birnbaum, 1968). However, the empirical study from Pelton (2002) shows that the estimation of the guessing parameters is unstable unless the parameters are made equal to a known or an unknown constant. Variants of IRT models have been developed to improve the modeling of guessing by modifying the parameters of IRT, for examples, difficulty plus guessing PL model (Kubinger and Draxler, 2006) and Ability-Based Guessing (1PL-AG) model (Boech and Leuven, 2006).

On the other hand, a psycho-technology approach to reducing the a priori guessing probability of an item (Kubinger *et al*., 2010) is proposed to discourage guessing. This approach involves two common ways of formatting the item response i.e., increases the number of

options but maintains single correct answer and increases the number of correct response options. Let the number of response options be k. For response format with single correct answer, the a priori guessing probability is $1/k$. By increasing the number of options but maintaining single correct answer, the a priori guessing probability will be less than $1/k$. In the case of increasing the number of correct response options, r, the calculation of the probability depends on whether the r is made known to the subjects. If r is known, the probability is related to the permutation of r correct response options and (k-r) distractors, which is given by $r!(k-r)!/k! = 1/{}^kC_r$. If the r is unknown, the a priori probability then amounts to $(1/2)^k$, which only depends on the number of response options. It can be seen that $1/k > 1/{}^kC_r > (1/2)^k$. In other words, the a priori guessing probability is the lowest if there are multiple correct response options and the number of correct response options is unknown. Kubinger *et al*. (2010) show that the difficulty parameter of the response format of 'two of the five response options are correct' is higher than 'one out of six response options is correct' when the r is made known.

Corresponding Author: Siow Hoo Leong, Department of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kota Samarahan, Sarawak, Malaysia Fax: +082677300 Tel: +6082677663

However, the finding focuses on the difficulties of the different response formats rather than the guessing effect of the subjects.

In this study, we consider a mixture of items with different number of multiple correct response options and therefore the priori guessing probabilities of these items are different. We propose a prior guessing parameter to be the moderator for the ability parameter in the psychometrics framework of IRT.

2. MATERIALS AND METHODS

2.1. IRT Model with Prior Guessing Parameter

We adopt the concept of testlet (Wainer and Kiely, 1987) to bundle the items with the same response option. The priori guessing probabilities for items within the same testlet are equal but different across testlets. We extend the notion of incorporating the testlet into IRT by Wang $et\ al.$ (2002) to propose a variant of testlet response theory to model the prior guessing effect by subjects.

Let the observed dichotomous responses of n subjects to m items be Y_{ij}, where $i = 1, 2,…, n$ and $j = 1, 2, …, m$. The item is scored as 1 if correct and 0 if not. In testlet response model, the conditional probability that subject i responses correctly to the item j ($Y_{ij} = 1$) is given by:

$$P\left(Y_{ij} = 1 \mid \theta_i, \gamma_{it(j)}\right)$$
$$= c_j + \left(1 - c_j\right) / \left(1 + \exp\left(\alpha_j\left(\theta_i - \beta_j + \gamma_{it(j)}\right)\right)\right) \qquad (1)$$

where, α_j, β_j and c_j are the discrimination, difficulty and guessing parameters respectively for item j, θ_i is the ability parameter of subjects i, $\gamma_{it(j)}$ is the testlet parameter accounts for the random effect of subject i across items that belong to the same testlet and t(.) is the function relates the belonging of items to the testlets, for example, t(1) = 2 means Item 1 belongs to Testlet 2. Each testlet parameter is assumed to follow normal distribution N(0, $\sigma^2_{t(j)}$) and represents a testlet effect through its own testlet specific variance, $\sigma^2_{t(j)}$. Procedures have been developed to estimate the variance of the testlet, Wainer $et\ al.$ (2007); Glas $et\ al.$ (2000) and Jiao $et\ al.$ (2013). Note that without the testlet effect, Equation (1) becomes:

$$P\left(Y_{ij} = 1 \mid \theta_i, \alpha_j, \beta_j, c_j\right)$$
$$= c_j + \left(1 - c_j\right) / \left(1 + \exp\left(\alpha_j\left(\theta_i - \beta_j\right)\right)\right) \qquad (2)$$

which is the 3PL-IRT model (Lord, 1980). The model assumes that an individual guesses item j correctly with probability c_j. If there is no guessing for all the items, $c_j = 0$ and Equation (2) is then reduced to 2PL-IRT model (Birnbaum, 1968) given by Equation (3):

$$P\left(Y_{ij} = 1 \mid \theta_i, \alpha_j, \beta_j\right) = 1 / \left(1 + \exp\left(-\alpha_j\left(\theta_i - \beta_j\right)\right)\right) \qquad (3)$$

In this study, we propose a 2PL-IRT model with testlet effect due to prior guessing. Consider m items with numbers of response options $k = k_1, k_2, …, k_d$ are bundled into d testlets $t_s = \{t_1, t_2…, t_d\}$ respectively. The priori guessing probabilities of these testlets are $\left(1/2\right)^{k_1}$, $\left(1/2\right)^{k_2}$,…, $\left(1/2\right)^{k_d}$ respectively. The proposed model is given by Equation (4):

$$P\left(Y_{ij} = 1 \mid \theta_i, \lambda_i\right) = 1 / \left(1 + \exp\left(-\alpha_j\left(\left(\theta_i + \lambda_{it(j)}\right) - \beta_j\right)\right)\right) \qquad (4)$$

where, $t(j) = t_s = \{t_1, t_2…, t_d\}$ for $j = 1, 2, …, m$, is considered as the testlet of items having the same number of multiple-correct response options and $\lambda_{it(j)}$ is the testlet effect due to guessing items with different priori guessing probabilities for subject i.

Let the number of items in the s-th testlet be m^s. The i-th subject has a d-dimensional response vectors $Y_i = (Y_{i1}, Y_{i2},…, Y_{id})$, where $Y_{is} = (Y_{i1s}, Y_{i2s},…, Y_{im^s s})'$ is the response vector at the s-th testlet and has a prior guessing effect vector $\lambda_i = (\lambda_{i1}, \lambda_{i2}, …, \lambda_{id})'$, where λ_{is} is the prior guessing effect at testlet t_s. The correlation among the prior guessing parameters measured on the same subject across d different testlets i.e., $\lambda_{i1}, \lambda_{i2},…, \lambda_{id}$, is expected to be higher than the prior guessing among differ-ent subjects within the same testlet i.e., $\lambda_{1s}, \lambda_{2s},…, \lambda_{ns}$. We consider this correlation structure of prior guessing as a moderator for the ability parameter. The distribution of the ability parameters is normal, we assume the prior guessing effect vectors are also come from a normal distribution, $\lambda_i \sim N(\mu_\lambda, \Sigma_\lambda)$, with mean structure, μ_λ and covariance, Σ_λ. The covariance matrix Σ_λ is considered unstructured and decomposed using Cholesky parameterization and becomes $\Sigma_\lambda = M_\lambda M_\lambda'$, where M_λ is the lower triangular matrix with positive diagonal elements and unrestricted elements below the diagonal. Higher value in Σ_λ indicates higher variation of the prior guessing effect for subject i across testlets with different multiple-correct response options and implies higher guessing in responding the items. Therefore, it can be used as a moderator for the ability parameter.

2.2. Simulations

Simulations are performed to study the inclusion of the proposed prior guessing parameter to moderate the ability parameter. Twenty items with multiple correct response options are considered. The items are bundled into 2 testlets with each consists of 10 items. The numbers of response options for the 2 testlets are $k_1 = 4$ and $k_2 = 5$ respectively but the numbers of correct response options are assumed unknown. Thus, the priori guessing probabilities of the 2 testlets are $(1/2)^4$ and $(1/2)^5$ respectively.

Response data of 9 groups of subjects which consists of 20 subjects each is generated. The subjects are considered to have 3 categories of ability i.e., low, average and high which respective ability parameters are -2.5, 0.5 and 2.5. The subjects respond to the testlet of items in 3 performance patterns i.e., poor, average and good. Difficulty parameters used to represent the performance patterns for all the 9 groups of subjects are shown in **Table 1**. The IRT models used to generate the data are considered to have constrained discrimination parameter, $\alpha_j = 1$ for $j = 1, 2, \ldots, m$. The first 3 groups are assumed to respond without guessing and 2PL-IRT model is used to generate the response data. For the rest of the 6 groups, 3PL-IRT model with guessing parameters equal to the priori guessing probabilities is used to generate the response data. Groups 4 and 5 are assumed to guess more than Groups 6 and 7. Groups 8 and 9 are having mixed performance pattern in both of the testlets. Our focus is only on the subjects with lower ability.

In this study, Bayesian estimation with Markov Chain Monte Carlo (MCMC) is used to estimate the parameters of the proposed response model. The MCMC method is not only provides a framework to experiment with new models (Kim and Bolt, 2007; Martin *et al.*, 2011), it is also more effective for heavy parameters IRT model (Baker, 1998; Azevedo *et al.*, 2012; Cho *et al.*, 2013). We adopt the prior distributions imposed on parameters $\theta_I \sim N(0, 1)$, $\alpha_j \sim N(0.8, 0.2^2)$, $\beta_j \sim N(0, 1)$ from Wainer *et al.* (2007) and $\Sigma_\lambda \sim$ gamma (0.5, 1), from Bradlow *et al.* (1999) which is proposed for testlet and to restrict the diagonal elements of M_λ to be positive. Random initial values are generated for the parameters. Since the number of iterations required for testlet parameters to converge is quite large (Sinharay, 2003; Sun *et al.*, 2012), we consider 10,000 iterations in our study. The Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002; Francois and Laval, 2011) developed as model selection method for Bayesian estimates of model parameters is used to compare the model fit of the pro-posed model and the benchmarked IRT model.

The simulations are performed using BUGS language implemented in OPENBUS version 3.2.1 (Lunn *et al.*, 2009) and the statistical programming environment R (RDCT, 2010) version 2.14.1.

3. RESULTS

The results of the proposed model are compared with 2PL-IRT model to evaluate the model fit and the performance of the prior guessing parameter in moderating the ability parameter.

Table 1. Simulated performance patterns and the difficulty parameters used for data generation

Group	Ability	Testlet	Performance patterns	Difficulty parameters
1	Low	1, $k_1 = 4$	Poor	1.5, 2
		2, $k_2 = 5$	Poor	1.5, 2
2	Average	1, $k_1 = 4$	Average	-0.5, 0.5
		2, $k_2 = 5$	Average	-0.5, 0.5
3	High	1, $k_1 = 4$	Good	-2, -1.5
		2, $k_2 = 5$	Good	-2, -1.5
4	Low	1, $k_1 = 4$	Poor	1.5, 2
		2, $k_2 = 5$	Good	-2, -1.5
5	Low	1, $k_1 = 4$	Average	-0.5, 0.5
		2, $k_2 = 5$	Good	-2, -1.5
6	Low	1, $k_1 = 4$	Good	-2, -1.5
		2, $k_2 = 5$	Average	-0.5, 0.5
7	Low	1, $k_1 = 4$	Good	-2, -1.5
		2, $k_2 = 5$	Poor	1.5, 2
8	Low	1, $k_1 = 4$	Mixed poor and good	1.5, -1.5
		2, $k_2 = 5$	Mixed poor and good	1.5, -1.5
9	Low	1, $k_1 = 4$	Mixed poor and good	1.5, -1.5
		2, $k_2 = 5$	Mixed good and poor	-1.5, 1.5

Table 2. Means of the estimated ability and prior guessing parameters for all the 9 groups

| Group | Proposed model | | 2PL-IRT model |
	Prior guessing	Ability	Ability
1	-1.60	-1.25	-1.36
2	-0.63	0.58	0.46
3	0.24	2.61	2.38
4	-0.44	-0.71	-0.32
5	-0.42	-0.50	-0.16
6	-1.52	-0.05	-0.37
7	-1.66	-0.10	-0.46
8	-0.93	-0.37	-0.36
9	-1.08	-0.42	-0.45

DIC selects the proposed model as a better model to fit the simulated data. The DICs for the proposed model and 2PL-IRT are 3325 and 3341 respectively. **Table 2** shows the means of the estimated ability and prior guessing parameters for all the 9 groups. The performance of the proposed prior guessing parameter is evaluated through three analyses as discuss below.

The first analysis focuses on the performance of the prior guessing parameter in the non-guessing groups of Groups 1, 2 and 3. The ability parameters of all the 3 groups estimated by the proposed model are higher than the 2PL-IRT model. The abilities estimated by the proposed model are adjusted by the prior guessing parameter by +0.11 (= -1.25-(-1.36)), +0.12 (= 0.58-0.46) and +0.23 (= 2.61-2.38) respectively for Groups 1, 2 and 3.

The second analysis evaluates the effect of the prior guessing parameter in moderating the abilities of subjects with different degrees of guessing. Groups 4 and 5 performed better in Testlet 2 which has lower priori guessing probability. These groups are assumed to have more guessing. The ability parameters of these groups estimated by the proposed model are lower than the 2PL-IRT model by -0.39 (= -0.71-(-0.32)) and -0.34 (= -0.50-(-0.16)) for Group 4 and 5 respectively. On the contrary, Groups 6 and 7 which have opposite performance pattern to Groups 4 and 5 are assumed to have less guessing. The estimated ability parameters of Group 6 and 7 are respectively +0.32 (= -0.05-(-0.37)) and +0.36 (= -0.10-(-0.46)) higher in the proposed model compare to the 2PL-IRT.

The third analysis evaluates the sensitivity of the prior guessing parameter across testlets with different priori guessing probabilities but with similar mixed performance pattern. The result shows that the ability parameters estimated by both models are very close.

The results also show that the scale of the prior guessing parameter for the subjects from the same ability level related to their performance pattern across testlets of items with different priori guessing probabilities. In the analysis, the low ability groups i.e., from Group 4 to Group 9 are considered. It can be seen that the prior guessing parameters are the lowest for Groups 6 and 7, which are -1.52 and -1.66 respectively. These two groups perform worse in the testlet with lower prior guessing probability. For groups that perform better in the testlet with lower prior guessing probability i.e., Groups 4 and 5, the prior guessing parameters are the highest, which are -0.44 and -0.42 respectively. For groups that with mixed performance pattern within testlet but similar across testlets i.e., Groups 8 and 9, the values of prior guessing parameters are between the range of two aforementioned clusters of groups.

4. DISCUSSION

This paper has described a psychometrics framework based on testlet response model to deal with guesing effect and shown that it measures subject's ability more reflectively than 2PL-IRT model. The proposed model introduces a prior guessing parameter, λ_{is}, which models the prior guessing effect of subject i at testlet t_s, in the testlet response model. The notion is adopted form Glas *et al.* (2000) where there are at least three mathematically isomorphic ways to include the testlet parameter in the IRT model. With the proposed prior guessing parameter, the logit of Equation (4) can actually be configured as $\alpha_j((\theta_i + \lambda_{it(j)}) - \beta_j)$ where $\lambda_{it(j)}$ as part of ability, or $\alpha_j(\theta_i + (\lambda_{it(j)} - \beta_j))$ where $\lambda_{it(j)}$ as part of difficulty, or $\alpha_j(\theta_i + \lambda_{it(j)} - \beta_j)$ where $\lambda_{it(j)}$ as an independent entity. However, the focus of this paper is the first case where prior guessing parameter as a part of ability is considered. The simulation results show that the proposed prior guessing parameter works well as a moderator for the ability parameter. The results from the first analysis imply that the prior guessing parameters from the proposed model merits ability for being not guessing. The second analysis result supports the first analysis and further implies that the prior guessing parameter merits ability of subjects with lower degree of guessing but penalize ability of subjects with higher degree of guessing. In terms of the scale of the prior guessing parameter, it is lower for subjects who show higher degree of guessing.

However, this model has some limitations. First, comparison between the third and first analysis results shows that the prior guessing parameter does not serve as a sensitive moderator in the case of similar performance pattern across testlets but mixed performance pattern within testlet. Second, the a priori guessing probability considered in this study is depending on the number of response options rather than the number of correct responses. In other words, the use of the design of multiple correct responses has not been utilised.

5. CONCLUSION

This study proposes a psychometrics framework of IRT to model the effect of having various priori guessing probabilities across items. The inclusion of the proposed prior guessing parameter in the 2PL-IRT model successfully serves as a moderator for the ability parameters. However, there are limitations on the model. The future works will be on two main scopes: the sensitivity of the prior guessing parameter towards the priori probability, number of items and number of testlets; how to model partial knowledge of the subject based on the design of multiple correct responses used in the proposed model.

6. ACKNOWLEDGEMENT

7. REFERENCES

Azevedo, C.L.N., D.F. Andradeb and J.P. Fox, 2012. A Bayesian generalized multiple group IRT model with model-fit assessment tools. Comput. Stat. Data Anal., 56: 4399-4412. DOI: 10.1016/j.csda.2012.03.017

Baker, F.B., 1998. An investigation of the item parameter recovery characteristics of a Gibbs sampling approach. Applied Psychol. Measur., 22: 153-169. DOI: 10.1177/01466216980222005

Birnbaum, A., 1968. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In: Statistical Theories of Mental Test Scores. Lord, F.M. and M.R. Novick (Eds.), Addison-Wesley, Reading, MA, ISBN-10: 0201043106, pp: 397-479.

Boech, P.D. and K.U. Leuven, 2006. IRT models for ability-based guessing. Applied Psychol. Measur., 30: 183-203. DOI: 10.1177/0146621605282773

Bradlow, E.T., H. Wainer and X. Wang, 1999. A Bayesian random effects model for testlets. Psychometrika, 64: 153-168. DOI: 10.1007/BF02294533

Cho, S.J., A.S. Cohen and S.H. Kim, 2013. Markov chain Monte Carlo estimation of a mixture item response theory model. J. Stat. Comput. Simulat., 83: 278-306. DOI: 10.1080/00949655.2011.603090

Francois, O. and G. Laval, 2011. Deviance information criteria for model selection in approximate Bayesian computation. Stat. Applied Genet. Mol. Biol., 10: 1-25. DOI: 10.2202/1544-6115.1678

Glas, G.A.W., H. Wainer and E.T. Bradlow, 2000. MML and EAP Estimation in Testlet-based Adaptive Testing. In: Computerized Adaptive Testing: Theory and Practice, van der Linden, W.J. and C.A.W. Glas (Eds.), Kluwer Academic Publishers, Netherlands, ISBN: 978-0792364252, pp: 271-287.

Jiao, H., S. Wang and W. He, 2013. Estimation methods for one-parameter testlet models. J. Educ. Measur., 50: 186-203. DOI: 10.1111/jedm.12010

Kim, J.S. and D.M. Bolt, 2007. Estimating item response theory models using Markov Chain Monte Carlo methods. Educ. Measur., 26: 38-51. DOI: 10.1111/j.1745-3992.2007.00107.x

Kubinger, K.D. and C. Draxler, 2006. A Comparison of the Rasch Model and Constrained Item Response Theory Models for Pertinent Psychological Test Data. In: Multivariate and Mixture Distribution Rasch Models-Extensions and Applications, Davier, M.V. and C.H. Carstensen (Eds.), Springer, New York, ISBN-10: 978-0387329161, pp: 293-309.

Kubinger, K.D., S. Holocher-Ertl, M. Reif, C. Hohensinn and M. Frebort, 2010. On minimizing guessing effects on multiple-choice items: Superiority of a two solutions and three distractors item format to a one solution and five distractors item format. Int. J. Select. Assess., 18: 111-115. DOI: 10.1111/j.1468-2389.2010.00493.x

Lord, F.M., 1980. Applications of Item Response Theory to Practical Testing Problems. 1st Edn., Erlbaum, Hillsdale, NJ., ISBN-10: 089859006X, pp: 274.

Lunn, D., D. Spiegelhalter, A. Thomas and N. Best, 2009. The BUGS project: Evolution, critique and future directions (with discussion). Stat. Med., 28: 3049-3082. DOI: 10.1002/sim.3680

Martin, A.D., K.M. Quinn and J.H. Park, 2011. MCMCpack: Markov chain monte carlo in R.J. Stat. Software, 42: 1-21.

Pelton, T.W., 2002. The accuracy of unidimensional measurement models in the presence of deviations for the underlying assumptions. Ph.D. Thesis, Brigham Young University, Provo, Utah, United States.

RDCT, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sinharay, S., 2003. Assessing Convergence of the Markov Chain Monte Carlo Algorithms: A Review. 1st Edn., Educational Testing Service, Princeton, NJ, ISBN-10: 08982236, pp: 59.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin and A.V.D. Linde, 2002. Bayesian measures of model complexity and fit. J. Royal Stat. Soc., 64: 583-616. DOI: 10.1111/1467-9868.00353

Sun, D., W.P. Sung and R. Chen, 2012. MCMC sampling statistical method to solve the optimization. App. Mech. Mater., 121-126: 937-941. DOI: 10.4028/www.scientific.net/AMM.121-126.937

Wainer, H. and G. Kiely, 1987. Item clusters and computerized adaptive testing: A case for testlets. J. Educ. Measur., 24: 185-202. DOI: 10.1111/j.1745-3984.1987.tb00274.x

Wainer, H., B. Bradlow and X. Wang, 2007. Testlet Response Theory and its Application. 1st Edn., Cambridge University Press, Cambridge, UK., ISBN-10: 9780521681261, pp: 280.

Wang, X., E.T. Bradlow and H. Wainer, 2002. A general Bayesian model for testlets: Theory and applications. Applied Psychol. Measur., 26: 109-128. DOI: 10.1177/0146621602026001007