A Bias Reduction Method for Ability Estimates in Adaptive Online IRT Testing Systems

Takenori Sakumura [*],   Hideo Hirose [†]

Abstract

To evaluate the ability of examinees, the item response theory (IRT) gives us an useful information. The IRT evaluates the examinee's ability and provides the item characteristics. Recently, we have configured the adaptive online testing using the IRT as one of the CBT (Computer Based Testing) methods, thereby we have been able to obtain the examinee's ability rating in short testing time. However, when the number of items is too small, we observed biases of estimates for the ability parameter. In this study, we have performed simulation studies under various conditions, and have found the magnitude of such biases. In addition, to circumvent the biases due to the Bayes procedure, we have proposed a simple method to reduce the biases.

Keywords: item response theory, adaptive test, item selection, item bank, ability estimate, bias

1 Introduction

In order to assess the ability of examinee, the item response theory (IRT) gives us valuable information [1–4]. A major difference between the traditional classical test theory (CTT) and the IRT, the IRT does not only assess the ability of examinees, but also evaluates the difficulties of items simultaneously. By using the characteristics of the IRT, we can implement the adaptive testing online (e.g. Computer Based Testing; CBT). Thus, we can expect to obtain the appropriate ability evaluation of examinees in short testing time [5]. However, when the number of items is rather small, the bias of the ability estimation value will begin to be observed [6–8]. In this study, we have performed simulation studies under various conditions, and have investigated the magnitude of such biases. Moreover, to circumvent the biases due to the Bayes procedure, we have proposed a simple method to reduce the biases.

---

[*] Department of Industrial and System Engineering, Chuo University, Tokyo, Japan
[†] Data Science Research Center, Hiroshima Institute of Technology, Hiroshima, Japan

2   Item Response Model

We define the probability for the correct answering response for item j by

$$
\begin{aligned}
P(\theta; a_j, b_j) &= Pr(U_j = 1; \theta, a_j, b_j) \\
&= \frac{1}{1 + \exp\{-Da_j(\theta - b_j)\}}.
\end{aligned}
$$

This is the two parameter logistic (2PL) model where $U_j \in \{0, 1\}$ denotes the binary response random variables; $U_j = 1$ means that the item j answered correctly and $U_j = 0$ means otherwise, $\theta \in (-\infty, \infty)$ is the ability parameter of the examinee, $a_j \in [0, \infty)$ and $b_j \in (-\infty, \infty)$ are the discriminating power and the difficulty on item j, respectively and D is a constant 1.702.

A set of items to be prepared for test, called an item bank, are denotes by $j = 1, \ldots, n$ and the identifiers of the items are denotes by $k = 1, \ldots, K$. Thus, j_k is the index of the item in the item bank. The likelihood function of a vector $u = \{u_{j_1}, u_{j_2} \ldots, u_{j_K}\}$ given θ is equal to the probability of a vector of observed responses u for an examinee of ability θ given by

$$
L = \prod_{j \in j_k} P(\theta; a_j, b_j)^{u_j} (1 - P(\theta; a_j, b_j))^{1 - u_j}.
$$

Note that we assume local independence. In a Bayesian approach, the posterior distribution of θ is expressed by

$$
g(\theta; u) = \frac{Lg(\theta)}{\int Lg(\theta) d\theta},
$$

where $g(\theta)$ is the prior distribution of θ. Here, we assume $g(\theta)$ is a standard normal distribution.

The ability estimators from the vector u are given by the maximum likelihood estimator $\hat{\theta}_{\mathrm{MLE}} = \arg\max_\theta L$ [9], and maximum a posteriori $\hat{\theta}_{\mathrm{MAP}} = \arg\max_\theta g(\theta; u)$ and expected a posteriori estimator $\hat{\theta}_{\mathrm{EAP}} = \int \theta g(\theta; u) d\theta$ in the Bayes theory [10]. As the other estimator, the weighted likelihood estimator WLE is given by Warm (1989) [11]. This ability parameter $\hat{\theta}_{\mathrm{WLE}} = \arg\max_\theta g_W(\theta; u)$ as

$$
g_W(\theta; u) = Lw(\theta; a_j, b_j),
$$

where $w(\theta; a_j, b_j)$ is the weight-function with respect to θ. In the one and two parameter models,

$$
\begin{aligned}
w(\theta; a_j, b_j) &= \sqrt{\sum_{j \in j_k} I(\theta; a_j, b_j)}, \\
I(\theta; a_j, b_j) &= \frac{P'(\theta; a_j, b_j)}{P(\theta; a_j, b_j)\{1 - P(\theta; a_j, b_j)\}}, \\
P'(\theta; a_j, b_j) &= \frac{\partial}{\partial \theta} P(\theta; a_j, b_j).
\end{aligned}
$$

Theses estimator can be separated in two classes of approaches to Frequentist approaches (MLE and WLE) and Bayesian approaches (MAP and EAP). Moreover, we also use the Markov chain Monte Carlo (MCMC) method with M-H algorithm with Gibbs sampler [12, 13].

3   Bias of the estimates for ability parameter under small samples

In this section, we examined the biases of the ability estimate $\hat{\theta}$ to the true ability value $\tilde{\theta}$ in accordance with the following:

$$bias = E[\hat{\theta} - \tilde{\theta}].$$

To estimate biases of the ability parameter, we have used a Monte Carlo simulations method under the various test conditions described below.

Item selection

We consider here the adaptive testing based on the responses to previous items using the Maximum-Information Criterion [14] (*adaptive*) and the non-adaptive testing (*non-adaptive*) in this paper. In the former case, *adaptive*, the next item for maximizing the item information $I(\theta; a_j, b_j)$ for θ is selected from the item pool. In the latter case, *non-adaptive*, the next item is randomly chosen from the item pool.

Item bank

The items prepared in advance to be administered as examination questions is called the item bank. In this simulations, we assumed two item banks which is *ideal* and *non-ideal*. The first item bank (*ideal*) has a perfect and infinite item bank having arbitrary difficulty parameters b_j, which $b_j \in \mathbb{R}$ for *adaptive* and $b_j \in [-3, 3]$ for *non-adaptive*. The second item bank (*non-ideal*) has infinite items at each of integer difficulty levels b_j, which $b_j \in \mathbb{Z}$ for *adaptive* and $b_j \in \{-2, -1, 0, 1, 2\}$ for *non-adaptive*. For example, when the item selection is adaptive and the item bank is ideal, an administered item is chosen from the item bank which has $b_j \in \mathbb{R}$ based on the Maximum-Information Criterion. The item discrimination parameter *a* for all items in the item bank is 1 or 2, that is, the correct answer probability for item *j* follows a one parameter logistic (1PL) model where all *a* is a common constant.

Stopping rule

The standard deviation $sd(\hat{\theta})$ is used as a convergence criterion to the adaptive test [15]. However, The bias of the estimated value is also important. Hence, we study the following two criteria.

1) the standard deviation of the ability estimate (standard error) is 0.3 or less and

2) the absolute value of the bias between the ability estimate and the true value is 0.1 or less.

In addition, the maximum number of administered items is $n = 30$.

Simulation Procedure

The simulation condition is as follows: Examinee i has a ability of the true ability value $\tilde{\theta} \in \{0, 1, 2, 3\}$. The item selection is $T \in \{adaptive, non\text{-}adaptive\}$. The item bank is $B \in \{ideal, non\text{-}ideal\}$. The ability estimate method is $M \in \{MLE, WLE, MAP, EAP, MCMC\}$ where *MLE* means the maximum likelihood estimate method, *WLE* means the weighted likelihood estimate method, *MAP* means the maximum a posteriori on the Bayesian estimate method, *EAP* means the expected a posteriori on the Bayesian estimate method, and *MCMC* means the Markov chain Monte Carlo method, which the number of chain is 1100 and burn-in is 100. The stopping rule is $S = \{\sigma_\theta \leq 0.3, |bias| \leq 0.1\}$ or the total number of items tackled is 30. The iteration cycle of the simulation is 10,000 to each condition. Table 1 shows the simulation conditions in this study.

The procedure of the simulation is as follows:

step 1: An examinee who has ability $\tilde{\theta}$ is administered the item j of the difficult parameter $b_j = 0$ as the first item.

step 2: The examinee is administered the item j which is chosen using the item selection method T based on the responses to the items up to the previous item.

step 3: If $P(\tilde{\theta}; a_j, b_j) > 0.5$, then the answer is correct, that is, $u_j = 1$, whereas if $P(\tilde{\theta}; a_j, b_j) \leq 0.5$, then the answer is incorrect, that is, $u_j = 0$. Based on the correct/incorrect pattern $(u_1, \ldots, u_j)$, we can estimate the ability value $\hat{\theta}$ using the estimation method M.

step 4: Repeat step 2 and 3 until satisfied the stopping rule S.

Note that the MLE and the WLE are not obtained when all the questions correct or incorrect. Therefore, we need a rule for such a situation. In this paper, the additional rule is as follows: if all the items is correct then $\hat{\theta} = 5$, and if all the items is incorrect then $\hat{\theta} = -5$.

We have repeated such simulations to check if the bias is observed in the testing methods mentioned above.

Simulation Results

Table 2 and Table 3 summarize the mean of the number of items required until the standard deviation of the ability estimate (standard error) becomes 0.3 or less

Table 1: Simulation conditions

The number of iteration is 10,000.

The maximum number of items is 30

The initial value $\theta_0 = 0$

The item selection T
    adaptive means the adaptive test.
    non-adaptive means the non-adaptive test.

The item bank B
    ideal means $b_j \in \mathbb{R}$ for *adaptive* or
                    $b_j \in [-3, 3]$ for *non-adaptive*.
    non-ideal means $b_j \in \mathbb{Z}$ for *adaptive* or
                     $b_j \in \{-2, -1, 0, 1, 2\}$ for *non-adaptive*.
    All $a = 1$ or 2.

The estimate method M
    MLE is the maximum likelihood estimation.
    WLE is the weighted likelihood estimation.
    MAP is the maximum a posteriori.
    EAP is the expected a posteriori.
    MCMC is the Markov chain Monte Carlo.

---

and the mean of the number of items required until the absolute value of the bias becomes 0.1 or less. Focusing on the standard error in the most appropriate conditions for $a = 1$, that is, the case of the item selection is *adaptive* and the item bank is *ideal*, we can confirm for $\tilde{\theta} = 1$ that MLE and WLE are required 18 items, MAP is required 15 items, EAP is required 17 items, and MCMC is required 16 items. In addition, the larger the true ability value, number of items is required a much more. MAP is smallest number of items for $\tilde{\theta} = 1$ and 2, whereas MLE is smallest number of items for $\tilde{\theta} = 3$. In this paper, we assume that the prior $g(\theta)$ is a standard normal distribution, therefore, the Bayesian estimators include MCMC can be estimated well around $\tilde{\theta} = 0$, however, it is much worse at $\tilde{\theta} = 3$. This tendency is clearer with the convergence criterion of $|bias| \leq 0.1$. In the bias criterion, we can confirm that 8 or more items are required in any estimation method. In particular, when true ability value $\tilde{\theta} = 3$, MLE and WLE are required at least 13 or more items, whereas MAP, EAP and MCMC are required at least 22 or more items. This means that the bias of the Bayesian estimator increases as $\tilde{\theta}$ increases.

    Figure 1 shows the mean bias of ability estimated value in each simulation conditions. The horizontal axis is the number of answered items, and the vertical axis is the bias. The ability estimate of MLE and WLE has positive biases, whereas the ability estimate of MAP, EAP and MCMC has negative biases. In this study, we have assumed that the $\hat{\theta} = 5$ or $\hat{\theta} = -5$ in the case of all the administered items are correct or all the administered items are incorrect, respectively. This is often observed in initial stages. Examinees who have high ability values are easily an-

swer correctly to all the administered items. Therefore, the positive biases appear for highly ability examinees. On the contrary, since the MAP estimation and EAP estimation are Bayesian estimations, we can obtain the estimated values even if all the administered items are correct. However, the estimated value is strongly affected by the prior distribution. Figure 1 shows that both MAP and EAP provide negative biases. These ability estimates are considered to be underestimated by the influence of the prior distribution. MCMC is also the same because it assumes prior distribution.

4   Correction of bias

In the previous section, we have shown the tendencies of biases under a variety of conditions. In this section, we consider a method for the correction of the biases.

Figure 2 is a part of a simulation result. The horizontal axis represents the true ability value $\tilde{\theta}$, the vertical axis represents the estimated ability value $\hat{\theta}$, and the 6 plot areas represents the number of administered items. In other words, we consider the relationship between the estimated ability and the true ability. Let the estimated ability value be $\hat{\theta}$, the true ability be $\tilde{\theta}$, and the number of items be m. The biases are following:

$$bias = h(\hat{\theta}; \tilde{\theta}, m)$$

In Figure 3, we assumed that a linear relationship between the ability estimate and the true ability value holds. That is,

$$h(\hat{\theta}; \tilde{\theta}, m) = \beta_{1,m}\tilde{\theta} + \beta_{2,m}\tilde{\theta}^2, \tag{1}$$

where $\beta_{k,m}, (k = 1, 2)$ is the regression coefficient for each number of items m. Further, the red line in Figure 2 shows a calculated regression equation. We describe the part of the regression coefficients of Equation (1) calculated as $\hat{\beta}_1$ and $\hat{\beta}_2$ in Table 4 and Table 5.

Then, the true ability value is determined as follows;

$$\tilde{\theta} = h^{-1}(\hat{\theta}, m; \hat{\beta}_{1,m}, \hat{\beta}_{2,m}). \tag{2}$$

Table 4, Table 5 and Figure 3 show the results of this correction as $\hat{\theta}^*$. The true ability value, the mean of the estimated ability parameters obtained in section 3, and the mean of the modified estimated ability parameters by Equation (2) with estimated regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are shown as $\tilde{\theta}$, $\hat{\theta}$, and $\hat{\theta}^*$, respectively. We can confirm that the modified estimated ability $\hat{\theta}^*$ becomes closely the true value $\tilde{\theta}$.

5   Conclusion

The adaptive testing is a useful test method in order to estimate the ability of examinee accurately. In practical applications, the item response theory played an

important role. We have shown the ability estimation methods, item selections, and stopping rules in simulation testing. To develop the optimum testing methods, various studies have been made in this study. If we want to perform the adaptive testing procedure using small number of problems, such as mathematics test, we have to take into account the bias of the ability estimate. In this study, we have done numerical simulations under a variety of conditions to investigate the bias of the ability value. As a result, we have found that the biases are observed by the influence of the prior distribution in MAP, EAP, and MCMC estimations. Therefore, we have found that the ability estimates are underestimated, and it is difficult to obtain the accurate ability value without some correction methods when the number of the items is less than 10. In addition, we have proposed a bias correction method using to a simple regression method.

Acknowledgment

References

[1] R. De Ayala, *The Theory and Practice of Item Response Theory*, ser. Methodology in the social sciences. Guilford Press, 2009.

[2] R. K. Hambleton and H. Swaminathan, *Item Response Theory: Principles and Applications*, 1985th ed. Boston : Hingham, MA, U.S.A: Springer, Dec. 1984.

[3] R. K. Hambleton, *Fundamentals of Item Response Theory*. Newbury Park, Calif: Sage Publications, Incorporated, Sep. 1991.

[4] W. J. van der Linden and R. K. Hambleton, Eds., *Handbook of Modern Item Response Theory*. New York, NY: Springer New York, 1997.

[5] D. J. Weiss, "Improving Measurement Quality and Efficiency with Adaptive Testing," *Applied Psychological Measurement*, vol. 6, no. 4, pp. 473–492, 1982.

[6] R. D. Bock and R. J. Mislevy, "Adaptive EAP Estimation of Ability in a Microcomputer Environment," *Applied Psychological Measurement*, vol. 6, no. 4, pp. 431–444, 1982.

[7] J. R. McBride, "Some Properties of a Bayesian Adaptive Ability Testing Strategy," *Applied Psychological Measurement*, vol. 1, no. 1, pp. 121–140, 1977.

[8] D. J. Weiss and J. R. McBride, "Bias and Information of Bayesian Adaptive Testing," *Applied Psychological Measurement*, vol. 8, no. 3, pp. 273–285, Jul. 1984.

[9] R. D. Bock and M. Aitkin, "Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm," *Psychometrika*, vol. 46, no. 4, pp. 443–459, Dec. 1981.

[10] F. B. Baker and S.-H. Kim, *Item response theory: Parameter estimation techniques*. CRC Press, 2004.

[11] T. A. Warm, "Weighted Likelihood Estimation of Ability in Item Response Theory," *Psychometrika*, vol. 54, no. 3, pp. 427–450, Sep. 1989.

[12] R. J. Patz and B. W. Junker, "A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models," *Journal of Educational and Behaviral Statistics*, vol. 24, no. 2, pp. 146–178, 1999.

[13] R. J. Patz and B. W. Junker, "Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses," *Journal of Educational and Behaviral Statistics*, vol. 24, no. 4, pp. 342–366, 1999.

[14] W. J. van der Linden and Wim J. *Elements of adaptive testing*. New York, NY: Springer, 2010.

[15] H. Wainer, N. J. Dorans, R. Flaugher, B. F. Green, and R. J. Mislevy, *Computerized Adaptive Testing: A Primer*, 2nd ed. Mahwah, N.J: Routledge, May 2000.

Table 2: Simulation results for *adaptive*; the iteration is 10,000, the number of items is 30, the initial value is $\theta_0 = 0$

Item selection	Item bank	True $\tilde{\theta}$	Estimation method	$a = 1$		$a = 2$	
				$sd(\theta) \leq 0.3$	$|bias| \leq 0.1$	$sd(\theta) \leq 0.3$	$|bias| \leq 0.1$
adaptive	*ideal*	1	MLE	18.51	11.73	7.11	8.50
			WLE	18.42	11.88	7.14	9.56
			MAP	15.93	13.06	5.69	9.85
			EAP	17.08	7.96	6.45	9.62
			MCMC	16.11	8.41	6.21	8.97
		2	MLE	19.18	12.60	7.76	9.42
			WLE	19.14	13.47	8.82	11.63
			MAP	17.77	17.29	7.58	11.46
			EAP	18.58	17.88	7.81	13.07
			MCMC	18.06	16.27	7.91	11.02
		3	MLE	19.77	13.04	8.40	10.21
			WLE	19.87	13.87	9.57	12.38
			MAP	20.59	23.37	11.16	17.44
			EAP	21.11	22.00	10.21	16.35
			MCMC	21.37	22.31	10.53	15.25
	non-ideal	1	MLE	22.43	14.54	8.22	10.97
			WLE	22.43	13.43	8.21	10.95
			MAP	20.32	11.20	8.01	13.39
			EAP	21.29	12.32	8.09	12.34
			MCMC	19.72	11.50	7.74	11.86
		2	MLE	23.30	15.44	9.23	11.86
			WLE	23.31	14.23	9.22	11.76
			MAP	21.19	18.04	8.68	10.78
			EAP	22.01	18.02	8.54	11.94
			MCMC	21.08	18.15	8.84	10.80
		3	MLE	24.24	16.33	10.22	12.52
			WLE	24.24	15.19	10.22	12.60
			MAP	22.24	22.82	9.96	16.95
			EAP	23.01	23.14	10.01	18.07
			MCMC	23.09	23.05	10.19	16.62

Table 3: Simulation results for *non-adaptive*; the iteration is 10,000, the number of items is 30, the initial value is $\theta_0 = 0$

Item selection	Item bank	True $\tilde{\theta}$	Estimation method	$a = 1$		$a = 2$	
				$sd(\theta) \leq 0.3$	$\|bias\| \leq 0.1$	$sd(\theta) \leq 0.3$	$\|bias\| \leq 0.1$
non-adaptive	*ideal*	1	MLE	29.15	15.35	15.34	14.82
			WLE	29.06	15.73	15.04	17.21
			MAP	27.55	14.56	13.68	14.91
			EAP	27.96	13.33	14.38	12.56
			MCMC	26.31	13.21	13.38	11.99
		2	MLE	29.98	20.70	25.66	23.90
			WLE	29.98	21.64	24.37	25.45
			MAP	29.84	23.42	20.82	20.15
			EAP	29.86	21.71	23.67	17.88
			MCMC	29.63	21.31	22.02	16.83
		3	MLE	30.00	29.66	29.98	30.00
			WLE	30.00	29.99	29.72	30.00
			MAP	30.00	30.00	29.32	30.00
			EAP	30.00	29.99	29.70	30.00
			MCMC	30.00	29.99	29.44	30.00
	non-ideal	1	MLE	29.82	14.84	17.85	12.54
			WLE	29.82	13.79	17.52	18.76
			MAP	29.18	15.24	16.13	19.27
			EAP	29.24	15.31	16.68	15.08
			MCMC	28.14	14.47	15.36	13.63
		2	MLE	29.99	19.62	21.20	18.55
			WLE	29.98	19.72	20.83	19.13
			MAP	29.87	22.64	18.64	15.64
			EAP	29.89	21.75	20.01	16.20
			MCMC	29.64	21.43	19.04	16.10
		3	MLE	30.00	27.21	29.87	30.00
			WLE	30.00	29.13	28.89	30.00
			MAP	30.00	29.90	28.15	30.00
			EAP	30.00	29.68	28.70	29.96
			MCMC	29.99	29.58	28.53	29.97

(a) MLE estimates

(b) WLE estimates

(c) MAP estimates

(d) EAP estimates

(e) MCMC estimates
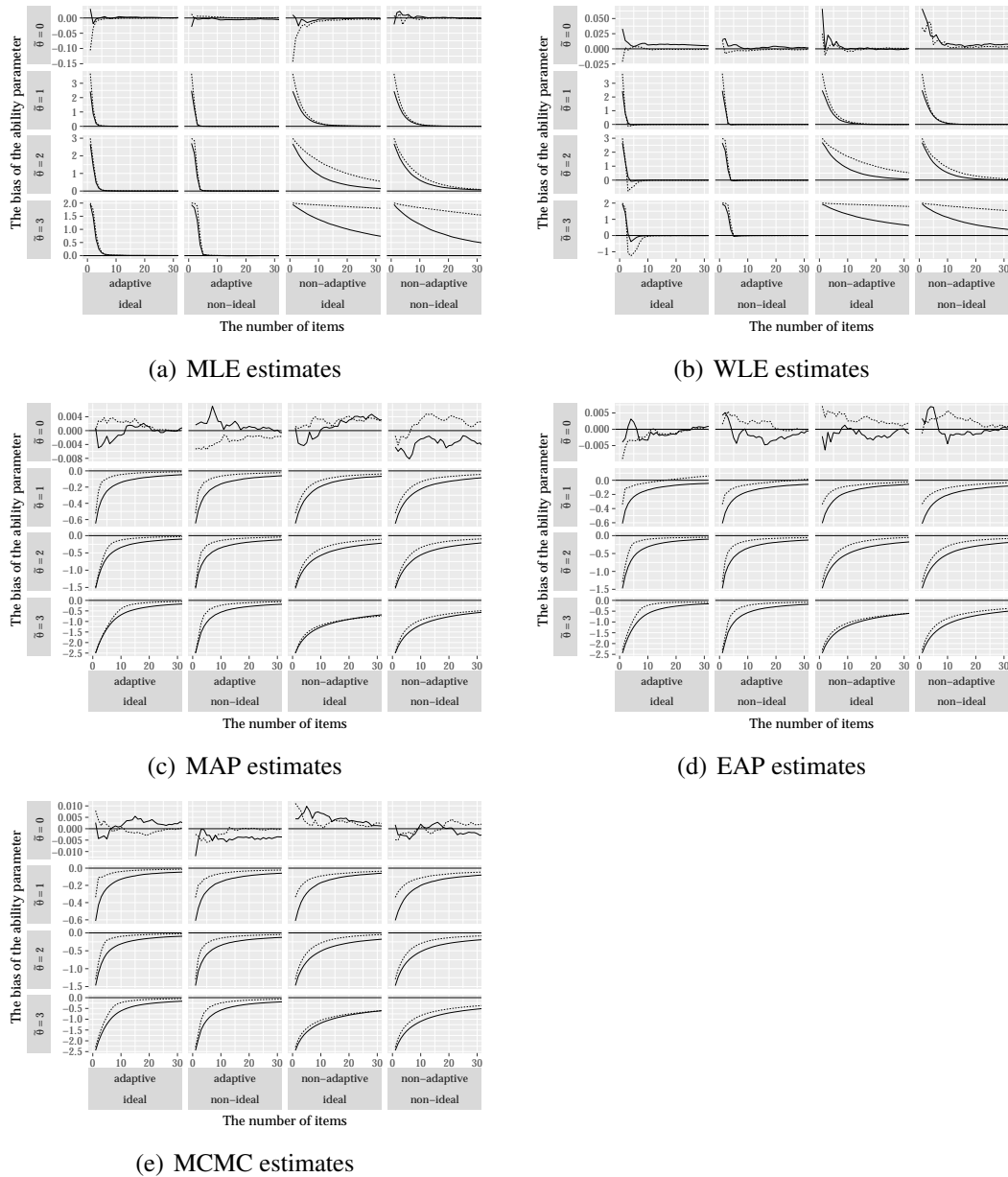
Figure 1: The mean bias of ability estimated value for each simulation conditions with $a = 1$ (*solid line*) and $a = 2$ (*dashed line*); Each column shows the conditions of the test and each row shows the true ability value in the 4x4 matrix plot.

Figure 2: Relationship of true ability value and ability estimated value of each answer number of items where $T = adaptive$, $B = ideal$, and $M = EAP$

Figure 3: A result of bias correction where $T = adaptive$, $B = ideal$, and $M = EAP$

Table 4: The regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ and the modified ability parameter $\hat{\theta}^*$ for each test conditions where $T = adaptive$, $B = ideal$, and $a = 1.0$.

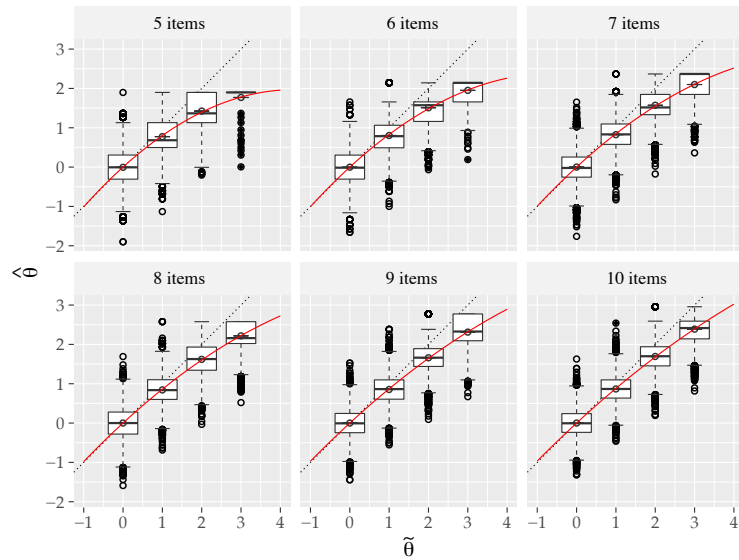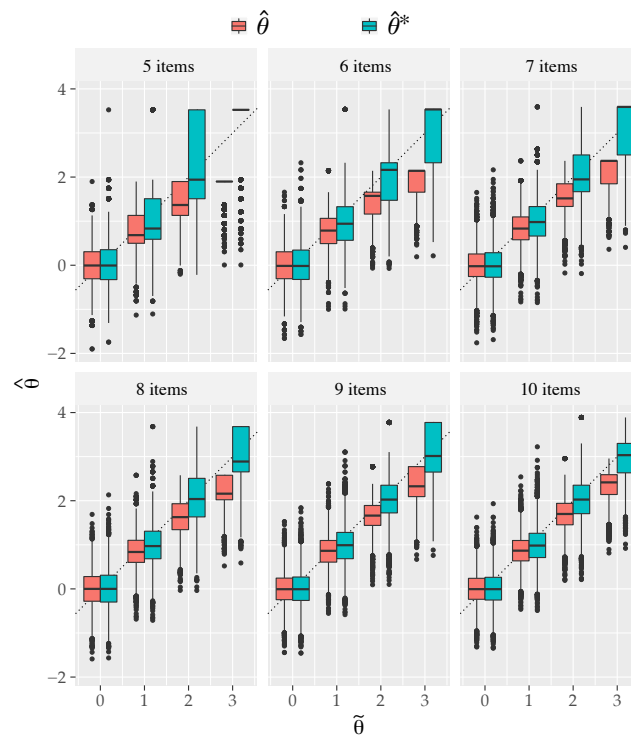| M	n	$\hat{\beta}_{1.}$	$\hat{\beta}_{2.}$	$\tilde{\theta}$	$\hat{\theta}$	$\hat{\theta}^*$	$\tilde{\theta}$	$\hat{\theta}$	$\hat{\theta}^*$	$\tilde{\theta}$	$\hat{\theta}$	$\hat{\theta}^*$
MLE	3	1.055	-0.075	1	0.973	1.080	2	1.819	2.100	3	2.487	3.004
	4	1.064	-0.030	1	1.028	1.014	2	2.015	2.034	3	2.919	3.031
	5	1.033	-0.008	1	1.028	1.006	2	2.033	2.002	3	3.030	3.006
	6	1.018	-0.002	1	1.020	1.005	2	2.025	1.997	3	3.039	3.002
	7	-	-	1	1.015	-	2	2.017	-	3	3.035	-
	8	1.009	-0.000	1	1.011	1.003	2	2.013	1.998	3	3.023	3.001
	9	-	-	1	1.008	-	2	2.010	-	3	3.020	-
	10	-	-	1	1.007	-	2	2.006	-	3	3.018	-
WLE	3	0.986	-0.091	1	0.894	1.065	2	1.611	1.887	3	2.142	2.409
	4	1.021	-0.041	1	0.973	1.023	2	1.886	2.052	3	2.694	3.059
	5	0.998	-0.010	1	0.984	1.001	2	1.960	2.011	3	2.901	3.006
	6	0.991	-0.001	1	0.989	0.999	2	1.979	2.002	3	2.962	3.000
	7	-	-	1	0.993	-	2	1.983	-	3	2.985	-
	8	-	-	1	0.993	-	2	1.989	-	3	2.988	-
	9	-	-	1	0.995	-	2	1.990	-	3	2.993	-
	10	-	-	1	0.995	-	2	1.993	-	3	2.993	-
MAP	3	0.771	-0.126	1	0.634	0.540	2	1.051	0.696	3	1.179	0.734
	4	0.846	-0.122	1	0.701	0.818	2	1.228	1.143	3	1.435	1.208
	5	0.887	-0.111	1	0.751	1.058	2	1.356	2.365	3	1.654	3.222
	6	0.905	-0.097	1	0.783	1.015	2	1.450	2.196	3	1.838	3.107
	7	0.912	-0.082	1	0.809	1.003	2	1.519	2.110	3	1.993	3.064
	8	0.915	-0.069	1	0.828	0.998	2	1.574	2.070	3	2.121	3.039
	9	0.918	-0.058	1	0.844	0.995	2	1.620	2.049	3	2.227	3.025
	10	0.922	-0.050	1	0.858	0.994	2	1.659	2.038	3	2.313	3.017
EAP	3	0.807	-0.125	1	0.667	0.564	2	1.127	0.728	3	1.287	0.756
	4	0.864	-0.115	1	0.725	0.849	2	1.295	1.204	3	1.554	1.269
	5	0.894	-0.100	1	0.766	1.030	2	1.416	2.261	3	1.775	3.136
	6	0.905	-0.083	1	0.795	1.005	2	1.502	2.142	3	1.955	3.073
	7	0.908	-0.068	1	0.817	0.996	2	1.564	2.085	3	2.101	3.043
	8	0.911	-0.057	1	0.835	0.992	2	1.616	2.058	3	2.218	3.026
	9	0.911	-0.046	1	0.849	0.992	2	1.653	2.040	3	2.313	3.016
	10	0.916	-0.039	1	0.863	0.992	2	1.689	2.030	3	2.391	3.010
MCMC	3	0.809	-0.126	1	0.668	0.815	2	1.130	1.689	3	1.289	2.428
	4	0.868	-0.115	1	0.725	0.990	2	1.302	2.051	3	1.557	2.871
	5	0.901	-0.102	1	0.767	1.026	2	1.425	2.276	3	1.774	3.142
	6	0.914	-0.086	1	0.798	1.001	2	1.512	2.158	3	1.954	3.081
	7	0.918	-0.072	1	0.823	0.996	2	1.573	2.092	3	2.102	3.047
	8	0.923	-0.060	1	0.843	0.994	2	1.623	2.060	3	2.218	3.031
	9	0.926	-0.051	1	0.860	0.994	2	1.663	2.042	3	2.313	3.020
	10	0.928	-0.043	1	0.872	0.995	2	1.693	2.030	3	2.388	3.013

T. Sakumura, H. Hirose

Table 5: The regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ and the modified ability parameter $\hat{\theta}^*$ for each test conditions where $T = adaptive$, $B = non\text{-}ideal$, and $a = 1.0$.

M	n	$\hat{\beta}_{1\cdot}$	$\hat{\beta}_{2\cdot}$	$\tilde{\theta}$	$\hat{\theta}$	$\hat{\theta}^*$	$\tilde{\theta}$	$\hat{\theta}$	$\hat{\theta}^*$	$\tilde{\theta}$	$\hat{\theta}$	$\hat{\theta}^*$
MLE	3	1.082	-0.210	1	0.892	0.529	2	1.295	0.731	3	1.393	0.744
	4	1.211	-0.143	1	1.014	0.905	2	1.904	1.563	3	2.312	1.762
	5	1.056	-0.028	1	1.007	0.990	2	2.021	2.040	3	2.909	3.008
	6	-	-	1	1.005	-	2	2.017	-	3	3.025	-
	7	1.005	-0.000	1	0.999	0.995	2	2.013	2.005	3	3.009	2.998
	8	1.003	-0.000	1	1.000	0.997	2	2.007	2.003	3	3.005	2.999
	9	1.003	-0.000	1	0.999	0.996	2	2.008	2.004	3	3.003	2.999
	10	1.002	-0.000	1	1.001	0.999	2	2.005	2.001	3	3.004	3.000
WLE	3	1.000	-0.190	1	0.827	0.470	2	1.213	0.647	3	1.318	0.677
	4	1.137	-0.130	1	0.965	0.899	2	1.799	1.515	3	2.217	1.717
	5	1.019	-0.027	1	0.979	0.998	2	1.942	2.032	3	2.808	3.012
	6	-	-	1	0.989	-	2	1.959	-	3	2.953	-
	7	-	-	1	0.992	-	2	1.976	-	3	2.973	-
	8	-	-	1	0.995	-	2	1.981	-	3	2.983	-
	9	-	-	1	0.995	-	2	1.985	-	3	2.985	-
	10	-	-	1	0.995	-	2	1.988	-	3	2.988	-
MAP	3	0.682	-0.062	1	0.610	1.071	2	1.127	2.218	3	1.486	3.132
	4	0.733	-0.048	1	0.681	1.039	2	1.276	2.069	3	1.763	3.109
	5	0.768	-0.038	1	0.723	1.012	2	1.392	2.046	3	1.960	3.035
	6	0.786	-0.027	1	0.761	1.017	2	1.461	2.012	3	2.116	3.022
	7	0.810	-0.023	1	0.785	1.006	2	1.529	2.014	3	2.220	3.012
	8	0.821	-0.016	1	0.807	1.008	2	1.574	2.003	3	2.315	3.008
	9	0.839	-0.015	1	0.822	1.002	2	1.619	2.007	3	2.379	3.005
	10	0.848	-0.012	1	0.839	1.006	2	1.648	2.000	3	2.441	3.005
EAP	3	0.695	-0.062	1	0.627	1.078	2	1.147	2.200	3	1.522	3.131
	4	0.736	-0.047	1	0.688	1.043	2	1.284	2.059	3	1.782	3.102
	5	0.773	-0.039	1	0.730	1.019	2	1.392	2.041	3	1.964	3.037
	6	0.792	-0.029	1	0.765	1.018	2	1.465	2.014	3	2.113	3.024
	7	0.814	-0.025	1	0.788	1.008	2	1.529	2.014	3	2.216	3.013
	8	0.826	-0.019	1	0.808	1.008	2	1.574	2.006	3	2.306	3.009
	9	0.844	-0.018	1	0.825	1.003	2	1.619	2.008	3	2.372	3.006
	10	0.852	-0.014	1	0.839	1.004	2	1.649	2.004	3	2.433	3.004
MCMC	3	0.700	-0.065	1	0.629	1.085	2	1.146	2.226	3	1.513	3.168
	4	0.752	-0.054	1	0.693	1.042	2	1.293	2.086	3	1.769	3.145
	5	0.784	-0.044	1	0.732	1.014	2	1.401	2.057	3	1.954	3.048
	6	0.805	-0.035	1	0.767	1.013	2	1.473	2.027	3	2.100	3.029
	7	0.828	-0.031	1	0.792	1.006	2	1.536	2.023	3	2.203	3.016
	8	0.841	-0.025	1	0.816	1.008	2	1.581	2.010	3	2.296	3.013
	9	0.853	-0.022	1	0.829	1.003	2	1.621	2.011	3	2.362	3.007
	10	0.861	-0.018	1	0.842	1.003	2	1.652	2.007	3	2.422	3.005