Int. J. Math. And Appl., 9(1)(2021), 41–50
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of *Mathematics And its Applications*

Effects of Item Parameter Distributions on Test Information Function

F. F. Ngwane[1,*]

1 Department of Mathematics, USC Salkehatchie, Allendale, SC 29810, United States.

Abstract: Test information function (TIF) plays an important role in designing a test. The TIF does not only depend on the logistic model such as the 1PL, 2PL and 3PL logistic models, but it also on the parameter distributions of the model. In this study we conducted simulations using the R statistical software on different examinee sizes and different exam length, and we compare the results when using uniform parameter distribution and normal parameter distribution. The findings of this study illustrate the following: First, when using item parameters that are all uniformly distributed the obtained test information function is significantly small compared to the test information function obtained when we use item parameters that all follow a normal distribution. Second, the test information function increases as the length of the exam increases when we use item parameters that are all normally distributed. Third, this study also suggest great flexibility in writing an exam as we can choose the item parameter distribution depending on the goals of the test.

Keywords: Item-Response Theory (IRT), Item Information Function, Test Information Function, one-, two-, and three-parameter models.

1. Introduction

Classical Test Theory (CTT) has been extended to Item Response Theory (IRT) which is a class of methods for modeling the relationship between the responses to the items or questions on a test, scale or questionnaire and the underlying latent trait(s) that the test is designed to measure (see Lord, 1975). IRT has four main assumptions, namely:

1) Monotonicity: – The probability of a correct response increases as the trait level; increases.

2) Unidimensionality: – There is one dominant latent trait being measured.

3) Local Independence: – Individual items responses are mutually independent at given ability levels

4) Invariance: – Item parameters can be estimated from any position on the item response curve.

IRT has been applied in many areas [8, 12]. Some commonly used Logistic Parametric (PL) models include the Rasch, !Pl, 2PL and 3PL models [2].

The Rasch model is given by

$$P_i(\theta) = \frac{1}{1+\mathrm{e}^{(\theta-b_i)}} \tag{1}$$

The 1PL, 2PL and 3PL models can be obtained from the general model below [1].

$$P_i(\theta) = c_i + \frac{1-c_i}{1+\mathrm{e}^{-1.7a_i(\theta-b_i)}} \tag{2}$$

* *E-mail: fifonge@gmail.com*

where b_i is the difficulty (or location) for item i, a_i is the discrimination (or scale or slope) for item i, c_i is the pseudo-guessing (or chance). When $a_i = 1$ for all i and $c_i = 0$ for all i, then we have the 1PL model. When $c_i = 0$ for all i then we have the 2PL model. Otherwise we have the 3PL model. θ indicates that the person abilities are modeled as a sample from a normal distribution.

In this study we seek to further existing studies on Item Information Function (IIF) and Test Information Function (TIF) by looking at how the item parameters affect the amount of information that a test can provide. Using the software R, we run some simulations on various item parameters and we calculate the test information. The simulations are done on long and short exams, large and small groups of examinees. The results are also presented graphically. We are interested in the 3PL model since the 1Pl and 2Pl models can be easily obtained from the 3PL model. The case of 4PL and other models will be studied in future.

2. Theoretical Framework

2.1. Item Information Function and Test Information Function

The probability that an examinee with a given ability level will answer an item correctly is given by the item response function. The higher the examinee's ability the more likely the examinee will answer the item correctly. We will consider dichotomous response data. An examinee with higher reading ability is more likely to get a reading item correct. Given the 1PL model, the item information is the probability of a correct response multiplied by the probability of an incorrect response. That is,

$$I(\theta) = p_i(\theta)q_i(\theta). \tag{3}$$

The item information function for the 2PL model is given by

$$I(\theta) = a_i^2 p_i(\theta)q_i(\theta). \tag{4}$$

while the item information function for the 3PL model is given by

$$I(\theta) = a_i^2 \frac{(p_i(\theta) - c_i)^2}{(1 - c_i)^2} \frac{q_i(\theta)}{p_i(\theta)}. \tag{5}$$

Test information function is the sum of all the item information functions and it is thus more informative than item information [9]. Item and test information functions are used to design tests for specific purpose [5].

Unlike the item information which is concentrated around ability levels that are close to its difficulty, the test information spreads the information over a wider ability range. The slope of the test information is determined by the item parameters. The latent ability and item parameters play an essential role in the test information function in predicting the expected score. If the discrimination parameters are positive, then the test information increases with ability and the higher the slope of the TIF, the more sensitive the expected score is to the differences in the ability. In 3PL models, the maximum is shifted from the difficulty.

An examinee's true ability at a given level can be estimated with greater precision if the test information is large. In this case, the ability estimate will be reasonably close to its true value. On the contrary, if the amount of information is small, then the ability cannot be estimated with precision, in which case the estimates will be widely scattered about the true ability.

The standard error of estimation (SE) is the reciprocal of the test information at a given trait level. Thus more information implies less error of measurement. The SE is given by:

$$\mathrm{SE}(\theta) = \frac{1}{\sqrt{I(\theta)}}. \tag{6}$$

The maximum information is given by:

$$\theta_{\text{max-info}} = b_i + \frac{1}{1.7a_i} \log\left(\frac{1+\sqrt{1+8c_i}}{2}\right). \tag{7}$$

2.2. Existing Results

In this section we present what has already been done. Item and test information functions have been compared for various parameter logistic models (PL), such as the 1PL, 2PL, 3PL [13]. Based on the data used in their study, the authors found that the 2PL model provided more information than the 1Pl and 3PL models. The authors also found that information provided by the 1PL models was not significantly different from the amount of information provided by the 3PL model.
The larger the test information, the smaller the measurement error. Some authors [10] have looked at various item parameters estimation errors in test development. This is important because if the parameter estimates are inaccurate then the test information will be inaccurate and so will the measurement error [10].

3. Motivation

This study is aimed at investigating how the item parameters affect the item information function and test information via simulations. Using long and short exams, large and small groups of examinees, we seek to examine how different item parameters affect the item information function and the test information function in the case of the Rasch model, 1PL, 2PL and 3PL models. Maximum information has been studied, [14].
The amount of information that an item or a test can provide is essential in writing a test. We seek to better determine and interpret the amount of information that an item or test can provide. Thus we ask what is the largest or maximum amount of information that a test can provide? Although the possible number of items to be included in a test has been studied [7], how do we select the best k items from a pool of $n+k$ good items? How do we get the most information from an exam, considering that we are limited by the number of questions that an exam can have, the examinee's ability and other factors involved in exam quality and exam scoring? How do we give each examinee a set of k good items than having each examinee take exactly the same exam that everyone else sees? To answer these important questions, we need to calculate the test information which is defined as the sum of the item information. This enables us to select the best questions to have on an exam. Item selection methods in the context of computer adaptive testing (CAT), Differential Item Functioning (DIF) and their applications have been studied, [3, 4, 6].

4. Simulations

We conducted simulations using small ($\theta = 100$), medium ($\theta = 500$) and large ($\theta = 1000$) number of examinees. Short ($n = 50$) and long exams ($n = 100$) were simulated. The descriptive statistics for the item parameters are in the tables below.

	Discrimination a_i	Difficulty b_i	Guessing c_i
minimum	0.5646526	-3.4111722	0.09070493
median	1.7474525	0.5847008	1.58859926
mean	1.9452878	0.1083172	1.80120526
standard deviation	0.8448833	2.1233612	1.02312880
maximum	3.4210322	3.3771706	3.43899366

Table 1. **Item parameters (uniformly distributed) for short ($n = 50$) exam**

	Discrimination a_i	Difficulty b_i	Guessing c_i
minimum	0.5425994	-3.46421640	0.007499111
median	2.2040057	-0.24586977	1.737994041
mean	2.1133071	-0.04623055	1.736667772
standard deviation	0.8589494	2.01724244	0.981234837
maximum	3.4695342	3.44869823	3.426066408

Table 2. **Item parameters (uniformly distributed) for long ($n = 100$) exam**

	Discrimination a_i	Difficulty b_i	Guessing c_i
minimum	-2.23200255	-2.802573384	-2.0002872
median	-0.11766173	0.024593358	-0.1144564
mean	-0.06789654	0.004206936	-0.1574632
standard deviation	0.91541166	1.136039067	0.9488562
maximum	1.53685513	2.331353708	2.6401935

Table 3. **Item parameters (normally distributed) for short ($n = 50$) exam**

	Discrimination a_i	Difficulty b_i	Guessing c_i
minimum	-2.79623064	-1.637280289	-2.42209343
median	0.00366203	-0.001881262	0.07811821
mean	-0.04135375	0.020533777	0.09630296
standard deviation	0.90011844	0.916995499	0.93649237
maximum	1.85718368	2.341201940	2.80956270

Table 4. **Item parameters (normally distributed) for long ($n = 100$) exam**

5. Results and Findings

The results obtained in this study are presented in Table 5 and Table 6. Using the uniformly distributed item parameters, we see that the test information function is large for short exams and small for long exams, regardless of the number of examinees. In Table 6, the item parameters follow a normal distribution. Here the smallest test information obtained is larger than any of the test information found in Table 5. This simulation study seems to suggest that we obtain much more test information when the item parameters are normally distributed. The results presented here also confirmed the inverse relationship between the information value and the standard error of measurement.

Number of examinees	Number of items	$\theta_{max-info}$	Max Test Information	Measurement Error
100	50	-0.6306926	1.358756	0.8578855
100	100	0.7839310	-1.559314	NaN
500	50	-0.6338395	1.357886	0.8581602
500	100	0.7548988	-1.554588	NaN
1000	50	-0.6306660	1.358760	0.8578840
1000	100	0.7608330	-1.554303	NaN

Table 5. **Test information and measurement error for uniform item parameters**

Number of examinees	Number of items	$\theta_{max-info}$	Max Test Information	Measurement Error
100	50	-0.5865101	15021.608	0.008159091
100	100	-0.1416374	1806.880	0.023525307
500	50	1.5780353	9242.426	0.010401764
500	100	0.1927761	1937.946	0.022715861
1000	50	-0.8510185	1378.527	0.026933471
1000	100	0.9541320	6973.245	0.011975194

Table 6. **Test information and measurement error for normal item parameters**

6. Conclusion and Discussion

The results obtained from this study is in accordance with existing results in the literature. We found that large information function implies small standard error of measurement. We found that as the guessing parameter decreases, the information function increases and thus the smaller standard error of measurement.

First, we consider the case when all the item parameters follow a uniform distribution. In this case we arrived at the following conclusions: (i) The item information function is large at maximum discrimination, minimum guessing, and average difficulty. (ii) The item information function increases for minimum discrimination, minimum guessing, and maximum difficulty. (iii) The item information function decreases for minimum discrimination, minimum guessing, and minimum difficulty. The length of the exam is not very significant in any of these three situations, see Figure 2 and Figure 4. This observation about the item information function is true when using either short or long exam.

Next we consider the case when all the item parameters follow a normal distribution. In this case we arrived at the following conclusions: (i) The smaller the discrimination, the smaller the item information function. (ii) The item information function is largest at maximum difficulty, minimum discrimination, and minimum guessing. (iii) Very little information is obtained when the guessing is maximum, see Figure 6 and Figure 8. The length of the exam is not significant.

The item information function decreases for minimum discrimination, minimum guessing, and minimum difficulty. The length of the exam is not very significant in any of these three situations. This study revealed three major findings: First, when using item parameters that are all uniformly distributed, the obtained test information function is significantly small (see Table 5) compared to the test information function obtained when we use item parameters that all follow a normal distribution (see Table 6). Second, the test information function increases as the length of the exam increases when we item parameters that are all normally distributed, see Figure 5 and Figure 7. This is not the case when the item parameters follow a uniform distribution, See Figure 1 and Figure 3. Third, the lower the ability and item difficulty, the higher the item information function. Also, the higher the ability and item difficulty, the higher the item information function. All these shows that there is great flexibility on writing an exam. We can choose the items on the test depending on the goals of the test.

Our future work will be to test our simulation results on some real life data sets and also investigate the possibility of any theory that may better explain the above observations.

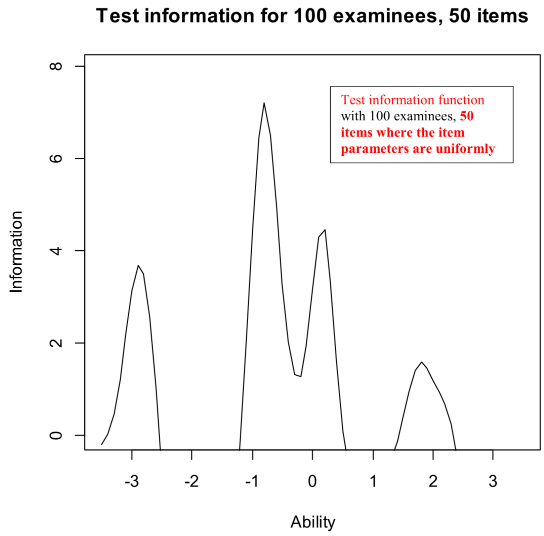

Figure 1. **Test information using 50 items with parameters that have a uniform distribution.**

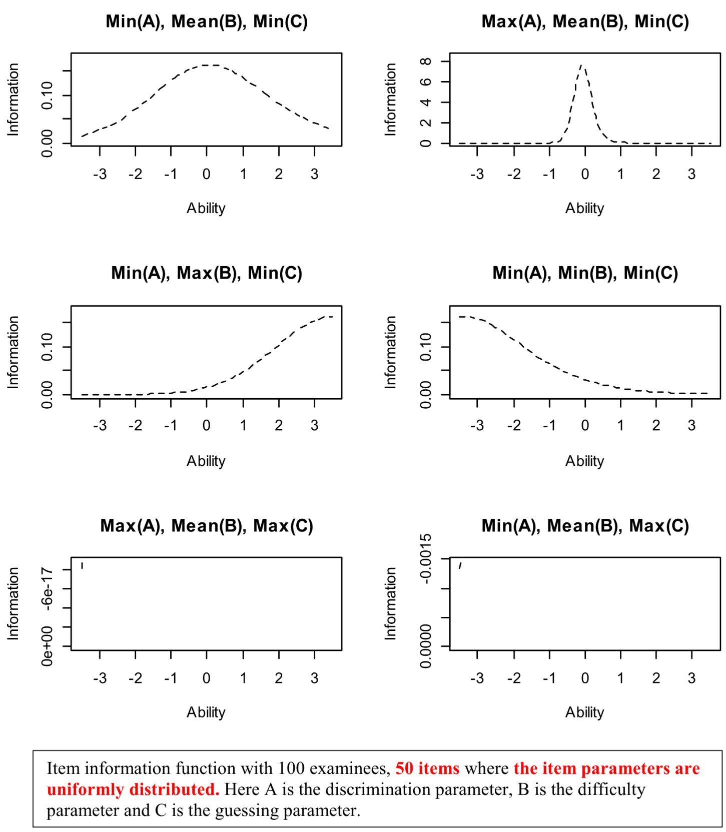

Figure 2. **Item information for selected items whose parameters have a uniform distribution.**

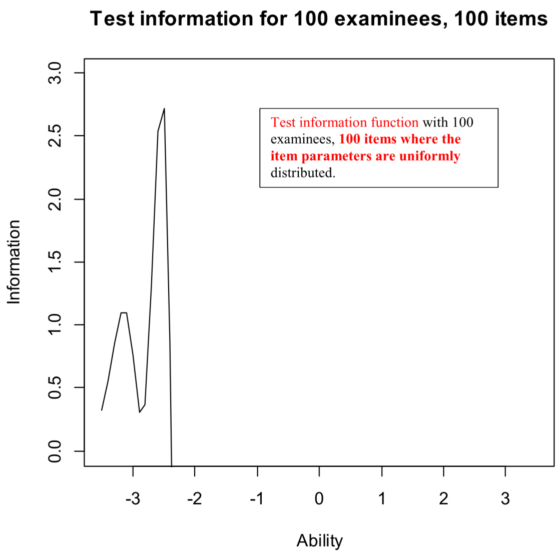

Figure 3. Test information using 100 items with parameters that have a uniform distribution.

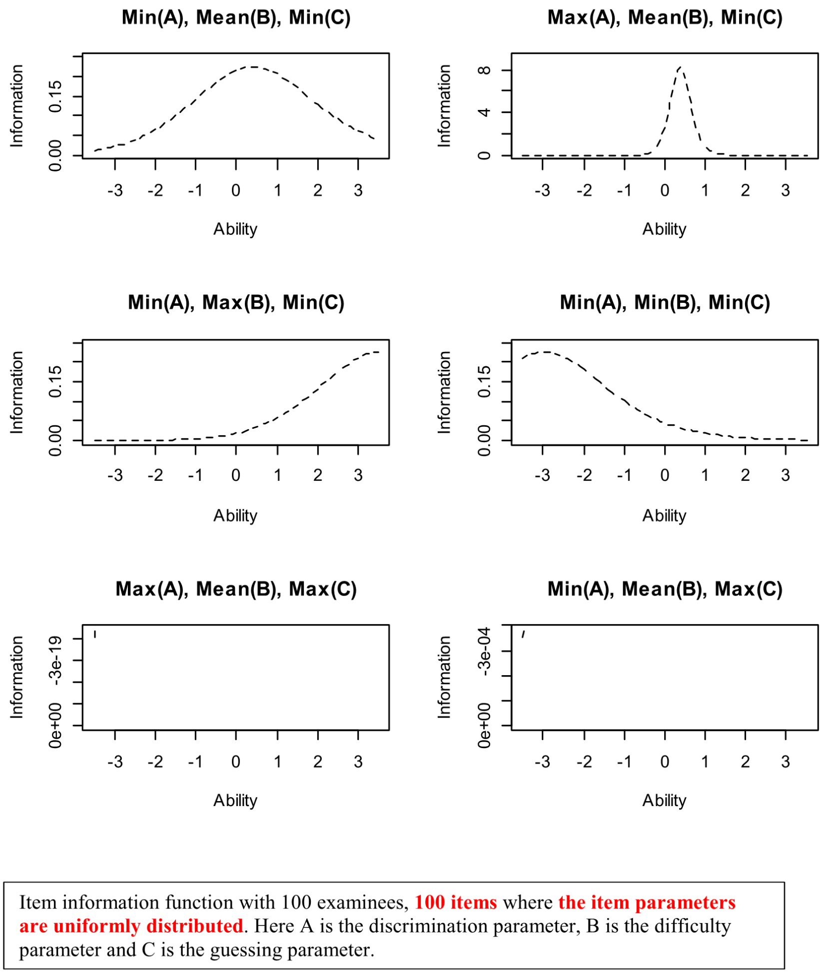

Figure 4. Item information for selected items whose parameters have a uniform distribution.

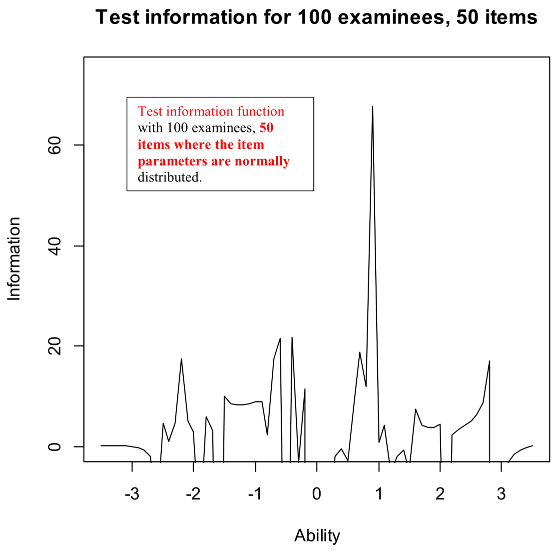

Figure 5. **Test information using 50 items with parameters that have a normal distribution.**

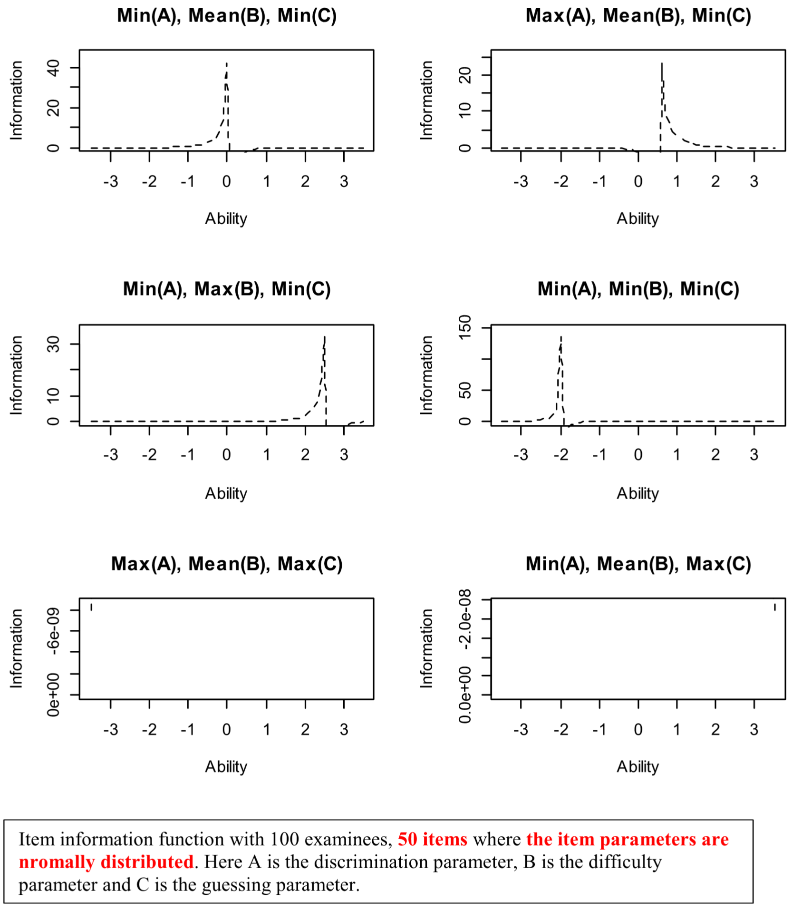

Figure 6. **Item information for selected items whose parameters have a normal distribution.**

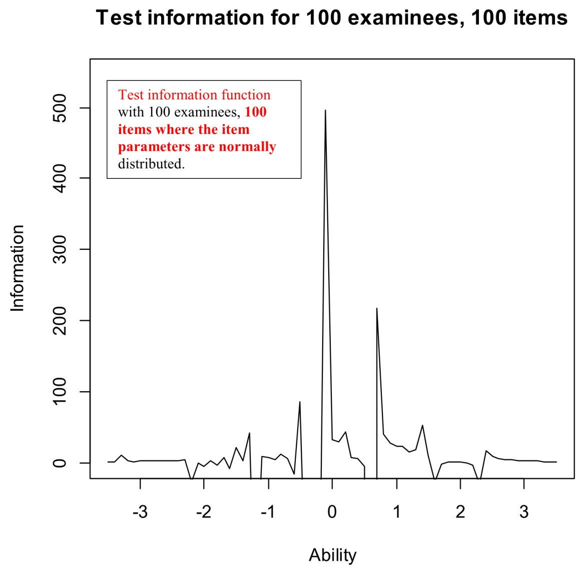

Figure 7. Test information using 100 items with parameters that have a normal distribution.

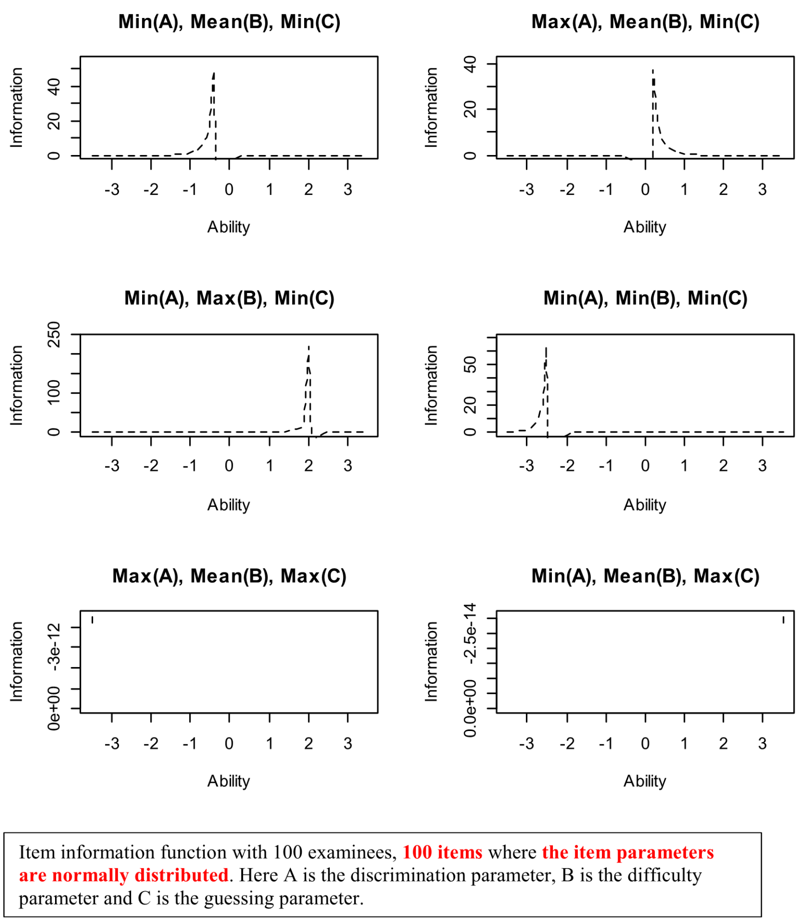

Figure 8. Item information for selected items whose parameters have a normal distribution.

References

[1] F. B. Baker and S. H. Kim, *Item Response Theory: Parameter Estimation Techniques*, 2nd Edition, CRC Press, Boca Raton, (2004).

[2] A. Brinbaum, *Some Latent Trait models and their use in inferring an examinee's ability*, In F.M Lord & M.R. Novick statistical theories of mental test score, Ma: Addison-Whesley, (1986).

[3] K. F. Cook, K. J. O'Malley and T. S. Roddey, *Dynamic assessment of health outcomes: time to let the CAT out of the bag?*, Health Services Research, 40(2005), 1694-1711.

[4] S. W. Choi and R. J. Swartz, *Comparison of CAT Item Selection Criteria for Polytomous Items*, (2009).

[5] S. M. Downing, *Item response theory: applications of modern test theory for assessments in medical education*, Medical Education, 37(2003), 739-745.

[6] M. C. Edwards, *An Introduction to Item Response Theory Using the Need for Cognition Scale*, Social and Personality Psychology Compass, 3(4)(2009), 507-529.

[7] S. E. Embretson and S. P. Reise, *Item response theory for psychologists*, Mahwah, NJ: Lawrence Erlbaum Assoc, (2000).

[8] R. K. Hambleton, *Principles and Selected applications of Item Response Theory*, Educational Measurement (3rd end), New York, Memillan, (1989), 147-200

[9] R. K. Hambleton and L. L. Cook, *Latent trait models and their use in the analysis of educational test data*, Journal of Educational Measurement, 14(2)(1977), 75-94.

[10] R. K. Hambleton, R. W. Jones and H. J. Rogers, *Influence of Item Parameter Estimation Errors in Test Development*, Journal of Educational Measurement, 30(2)(1993), 143-155.

[11] F. M. Lord, *The "ability" Scale in item characteristic curve theory*, Psychometrika, 40(1975), 205-217.

[12] F. M. Lord, *Applications of item Response theory to practical Testing Problems*, Hillsdale, N.G.: Lawrence Ertaun, 14(1980), 117-138.

[13] A. Moghadamzadeh, K. Salehi and E. Khodaie, *A Comparison the Information Functions of the Item and Test in One, Two and Three Parametric Model of the Item Response Theory (IRT)*, Procedia - Social and Behavioral Sciences, 29(2011), 1359-1367.

[14] W. J. J. Veerkamp, P. F. Martijn and M. P. F. Berger, *Optimal Item Discrimination and Maximum Information for Logistic IRT Models*, Applied Psychological Measurement, 23(1)(1999), 31-40.