E – 29

Assembling the Mathematics Test Using the Value of Information Functions

Heri Retnawati

Mathematics Department, Mathematics and Science Faculty,
Yogyakarta State University
retnawati_heriuny@yahoo.co.id

Abstract

Assembling the mathematics test from the items have been provided in the item bank requires a special way. This method is related to the chosen theoretical approach in developing test that appropriate with its objectives. In this paper is presented how to assemble the test item response theory approach. Each item in the item bank administered its parameters, i.e. the difficulty index, discrimination index, or pseudo guessing. With these parameters, the value of information function can be computed, and based on the value, the standard error of measurement (SEM) can be estimated. The steps of assembling the test using the item bank are determines the desired SEM, specify the target of value of information function (VIF), drawing the target of VIF, inserting items from item bank that determine the achievement of the target of VIF, these items are then compiled into a test, using criteria of a good test preparation, for example, sorted from the easiest to the most difficult.
Keyword: assembling test, value of information function (VIF)

Introduction

One application of the item response theory is assembling a test. The test can be assembled using an item bank, which has been developed in accordance with the purpose of the test. For example, to held the final examination of mathematics lesson in Junior high school, required an item bank or a set of mathematics items or a pool of items and the items have known its characteristics. The question bank is not only a storage system that organizes, but also system how to select items for use, and the elimination of items that can't be used anymore. The use of these items adapted to the purpose and objectives of development of the test. Using the unidimensional item response theory, item banks can be developed by using three models, namely the one parameter logistic model (1PL), the two parameters logistic model (2PL) and the three 3 parameters logistic model (3PL). The item parameters are the difficulty index (1PL), the difficulty index and the discrimination index (2PL), the difficulty index, the discrimination index, and pseudo-guessing parameter (3PL). Using these parameters, the ability of the testee (θ) can be estimated after doing a series of test items, and the relationship with the probability of correctly answering item parameters 3PL models presented mathematically in Equation 1 (Hambleton, Swaminathan, and Rogers, 1991:17, Hambleton, and Swaminathan, 1985:49, Vander Linden and Hambleton, 1997:13).

$$P_i(\theta) = c_i + (1-c_i)\ \frac{e^{a_i(\theta-b_i)}}{1 + e^{a_i(\theta-b_i)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (1)$$

In the equation, a_i is an item discrimination index, b is the difficulty index, and c is pseudo guessing parameter. The 1PL model is a special case of the 3PL with a =1 and c=0, and the 2PL model is a special case of with c=0. As an illustration, Figure 1 is item characteristic curves with 3 items, item 1(a = 1, b=0.5, c=0), item 2 (a = 0.5, b=0.5, c=0) and item 3 (a=0.5, b=0.5, c=0.2).
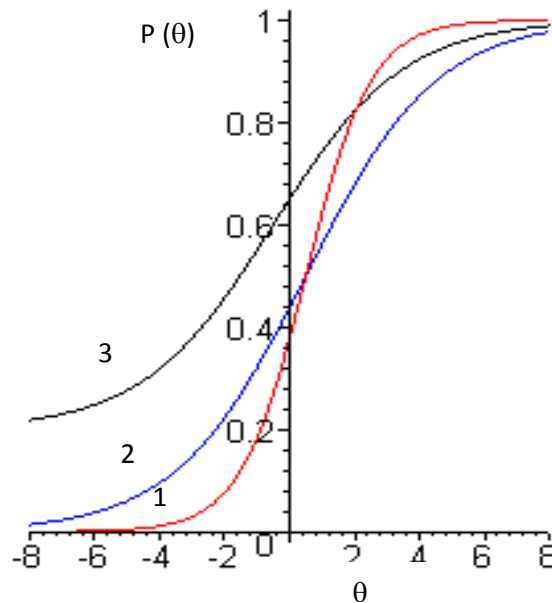

Figure 1. The item characteristic curves in 3PL models, with item 1(a = 1, b=0.5, c=0), item 2 (a=0.5, b=0.5, c=0) and item 3 (a=0.5, b=0.5, c=0.2)

The value of the ability of testee (θ) located between -4 and +4, according to the domain of normal distribution. This statement is an underlying the assumption that a large value of b_i. In theory, the value of b is between -∼ and +∼. An item said to be good b parameter if the value is between -2 and +2 (Hambleton and Swaminathan, 1985:107). If the value of b_i approaching-2, then the item difficulty index is very low, where as if the value b_i +2 then the item difficulty index is very high for a group of test takers.

The a_i parameter is a discrimination index of the i-th item. On the characteristic curves, a_i is the slope of the curve at point b_i on a scale of certain abilities. Because a_i is a slope, the greater the slope obtained, the greater the discrimination of the item. In theory, the value of a_i is located between and -∼to +∼. In good items, the a_i parameters have positive relationship with the performance to estimate the ability and a_i located between 0 and 2 (Hambleton and Swaminathan, 1985:37).

The probability of the correct answer when the ability is -∼ denoted by c_i, which is called the pseudo guessing. This parameter gives a possibility of non-zero lower asymptote in the item characteristic curve (ICC). This parameter describes the probability of testee with a low ability to answer correctly on an item that has a difficulty index that is not in accordance with the participant's ability. The magnitude c_i assumed less than the amount that would result if the testee answered randomly to an item. The value of c_i is between 0 and 1. In a good item, the value of c_i is not more than 1/k, with k is the number of options (Hullin,1983:36).

Related to the characteristics of the items and item bank, only good items with good characteristics that can be stored in the item bank. For item bank developed using the item response theory, the difficulty index can span from easy to difficult, discrimination index between 0-2, and for multiple-choice items, c_i does not exceed 1/k with k number of choices. An item from a bank can be used with assembled into a test device that corresponds to the bank designation of tests.

There are several ways to select items from an item bank to be assembled into a test. The first way is the most commonly used, with due regard to the content matter. By looking at the indicator of items, it can be assembled a test. The second way is focused on the content and level of difficulty. This method considers the difficulty level of information available in the item bank. The third way to utilize the value of information function (VIF). This method is not commonly used, because the item response theory approach. Related primarily to the third way, in this paper discussed the utilization of VIF to assemble the test accompanied by simulation.

Discussion

Item information functions is a function to describe the strength of an item on the test and declared the strength or contributions of items in uncovering latent ability (latent trait) as measured by those tests. Using information function, the item that suitable with the ability of testee. It will help in the selection item. Mathematically, the item information function is defined as follows.

$$I_i(\theta) = \frac{[P_i^{'}(\theta)]^2}{P_i(\theta)Q_i(\theta)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

where:

i   : 1,2,3, ..., n

$I_i(\theta)$: the item information function of the i-th item

$P_i(\theta)$: probability of testee that have ability θ to answer correctly i-th item

$P'_i(\theta)$: the derivative of the function $P_i(\theta)$ to θ

$Q_i(\theta)$: probability of testee that have ability θ to answer uncorrectly i-th item

The item information function for the three-parameter logistic model is expressed by Birnbaum (Hambleton & Swaminathan, 1985:107) in the following equation.

$$I_i(\theta) = \frac{2,89a_i^2(1-c_i)}{\left[(c_i + \exp(Da_i(\theta-b_i)))\right]\left[1+\exp(-Da_i(\theta-b_i))\right]^2} \dots\dots..(3)$$

where a_i, b_i, c_i are item parameters, e natural number (approaching 2,718), and D=1,7.

Based on the information function equation above, then the information functions satisfy the properties: (1) the item response logistic model, the function item information reach a maximum value when bi approaching θ, in the three-parameter logistic model maximum value is achieved when θ is located slightly above b_i and the index of pseudo guessing decreases; (2) increasing the overall information function happened if the discrimination index increases. De Gruijter & Vander Camp (2005: 118) states that the value of item information functions and also the value of the test information function depends on the latent ability.

The test information function is the sum of the information functions of the items (Hambleton & Swaminathan, 1985: 94). In this regard, the value of the test information function will be high if the every item making up the test has a high information functions as well. The test information function (I (θ)) can be defined mathematically as follows.

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \quad \text{(4)}$$

The Indexes of item parameters and the ability of testee are the result of estimation. Because the estimation results, the truth is probabilistic and had a measurement error. In item response theory, measurement error standardized (called Standard Error of Measurement, SEM) is closely related tothe function information. SEM has inverse quadratic relationship with information function, the greater the information function, the smaller SEM or vice versa (Hambleton, Swaminathan, &Rogers, 1991, 94). If the value of the information function is expressed by and the estimated value of SEM revealed by SEM(θ), then the relationship between the two both $I_i(\theta)$ and SEM, according to Hambleton, Swaminathan, & Rogers (1991: 94) is expressed as follows.

$$SEM\,\overset{\wedge}{(\theta)} = \frac{1}{\sqrt{I(\theta)}} \quad \text{(5)}$$

As an illustration, a graph in Figure 2 presented the value of item information function and standard error of measurement of an item with parameters a=2, b=-0.5, and c=0.1.
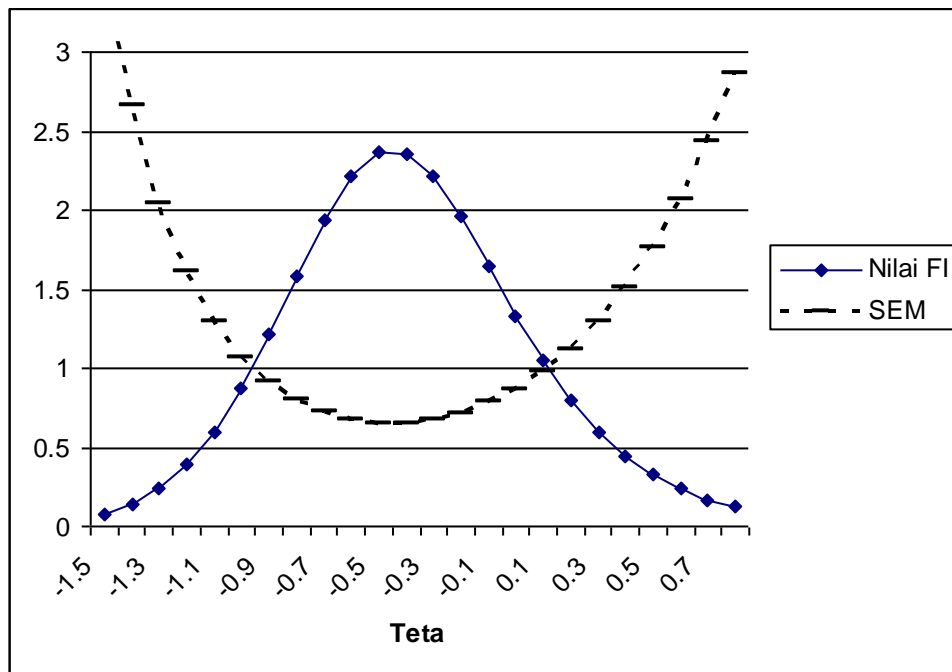

Figure 2
Value of Item Information Functions and Standard Error Measurement of an Item
with parameter a=2, b=-0.5, and c=0.1

International Seminar on Innovation in Mathematics and Mathematics Education
1st ISIM-MED 2014 Department of Mathematics Education,Yogyakarta State University,Yogyakarta,
November 26-30, 2014

EP - 248

Interpretation of Figure 2 is as follows. The value of the information function of the ability-~rose, reaching a maximum value, then decreases to +~. While the measurement error on the contrary, decreases, reaches a minimum value, then increases again. Both graphs of these functions meet on a scale of -0.9 and +0.2 ability. Between the set of abilities, the item has a value of information function higher than the measurement error. Conversely, when the scale is less than -0.9 and the ability is more than +0.2, the item has measurement error more the value of information function.

Similarly on the test, the value of the information function of the test is the sum of it VIF's items. As illustrasion in Figure3, which presents the value of the test information function of a junior high school mathematics national examination with 30 items. This means the has VIF higher than SEM in the range of-1.7 to +1.8. Thus, the test is suitable for testing if the testee is on the range abilities. If the scale is less than-1.7 and the ability is more than +1.8 measurement error is greater than the value of the information function.
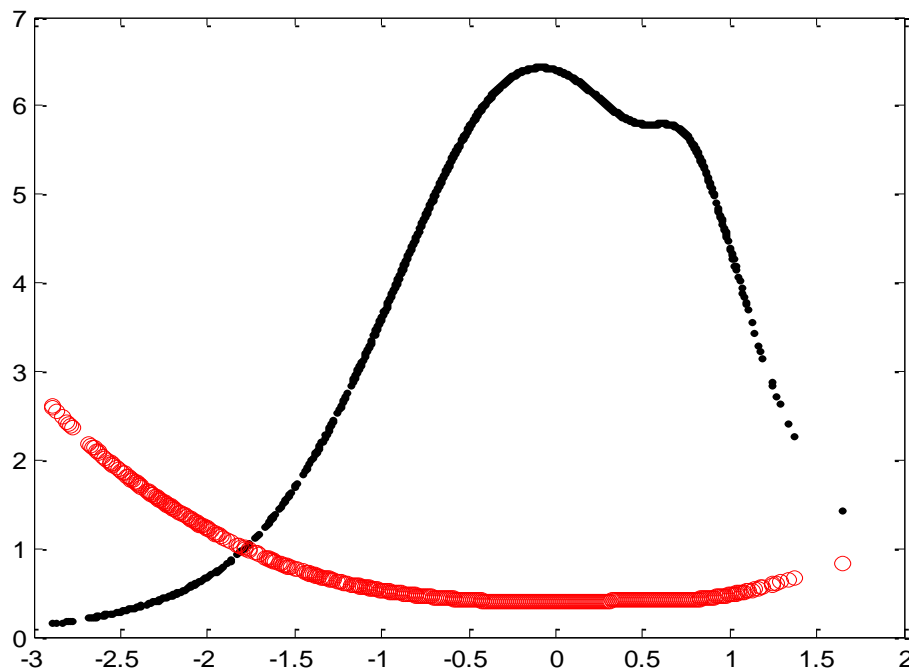

Figure 3
The Value of Information Function of Mathematics Test

Along with the development of science and technology, and related with VIF and SEM, the assembly ability of the test adjusted the testee' ability and formatted in a computer not in paper and pencil test. It is called as computerized adaptive testing (CAT). This is done to obtain the value of the test information function as much as possible and to minimize the measurement error. Thus, the test can be assembled according to the ability of testee, for example, for high ability group, medium, and low testee. This anticipates the case of participants with low capacity and are doing test for

medium group, it can obtain higher measurement error than the information. It needs to be assembled a test to suit specific testee' abilities.

In this test assembly, the initial capital required is the availability of item bank. The item bank is management system that both stores the good items and provide a mechanism how to use the items. In order to user can use the item freely, the item bank should save quite a lot of good items with various levels of difficulty index.

The steps of assembling the device by using the item bank are follows.

1. Determine the desired SEM

    For example, an expert who want to use the test measurement error of 0.316228.

2. Determine the target VIF and the ability of testee

    Using equation 5, it can be calculated the target of VIF that is 10, and it will be used for testee with the ability from -2.5 to -0.8.

3. Drawing VIF targets

    Furthermore, the VIF is described in Cartesian coordinates, with the horizontal axis and vertical axis capability function value information.

4. Incorporating the items of banks that determine the achievement of the VIF target

    Items that supporting VIF selected when the number of VIF approached the VIF target.

5. The selected items in step 4 then compiled into a test, using the criteria of a good test preparation, such assorted from the easiest to the most difficult.

For the simulation to use VIF for assembling a test, it is generated 200 item parameters under the3PLmodels, in item bank with the difficulty index has uniform distributiron in the range (-3, +3), discrimination index in the range (0,2) and the pseudo guessing parameters are than 0.25 using WINGEN. Furthermore, basic on the steps to assemble the test of the item bank, simulated 3 test for group in low, medium, and high level ability.

At the low group, using 20 items and the mean difficulty index is -1.84 obtained VIF close to the target VIF. The test is suitable for participants with the ability scale from -4.0 to -0.3 (-4.0 , -0.3). The results further are illustrated in Figure 4.
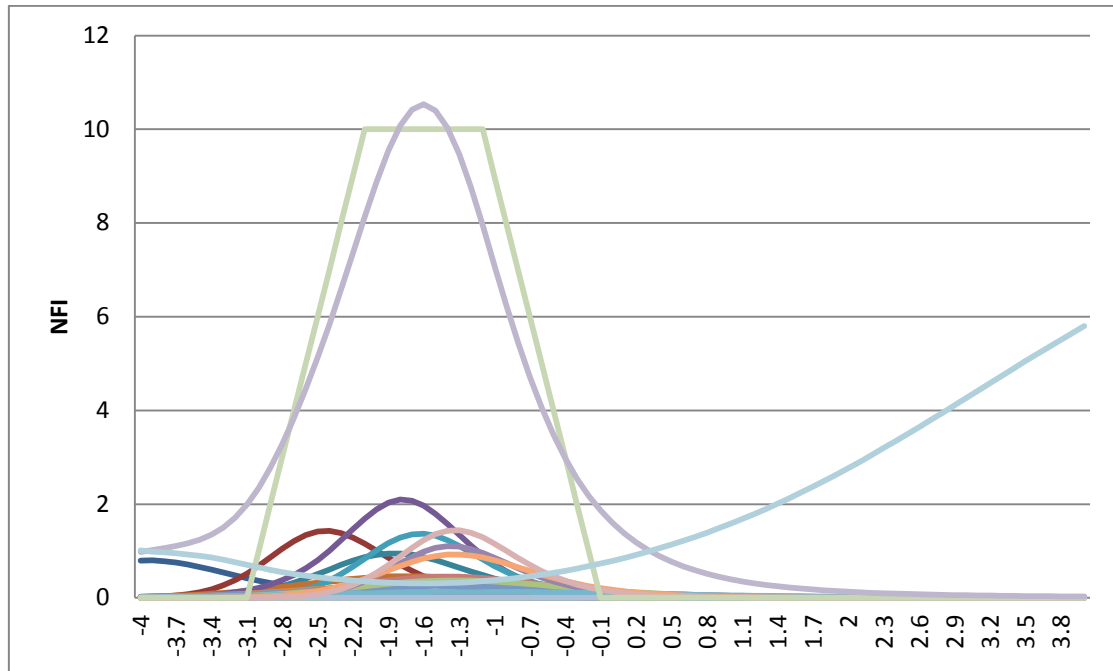
International Seminar on Innovation in Mathematics and Mathematics Education
1ˢᵗ ISIM-MED 2014 Department of Mathematics Education,Yogyakarta State University,Yogyakarta, November 26-30, 2014

EP - 250

Figure 4

Value ofthe target VIF and test VIF

and the VIF items that composed for low ability testee

For the medium ability group, using 18 items and the mean difficulty index is +0.315, obtained VIF close to the target VIF. The test is suitable for participants with the ability scale from (-1.2, +2.4). The results further are illustrated in Figure 5. For the high ability group, using 18 items and the mean difficulty index is +1.84, obtained VIF close to the target VIF. The test is suitable for participants with the ability scale from (-0.4, +4.0). The results further are illustrated in Figure 6.
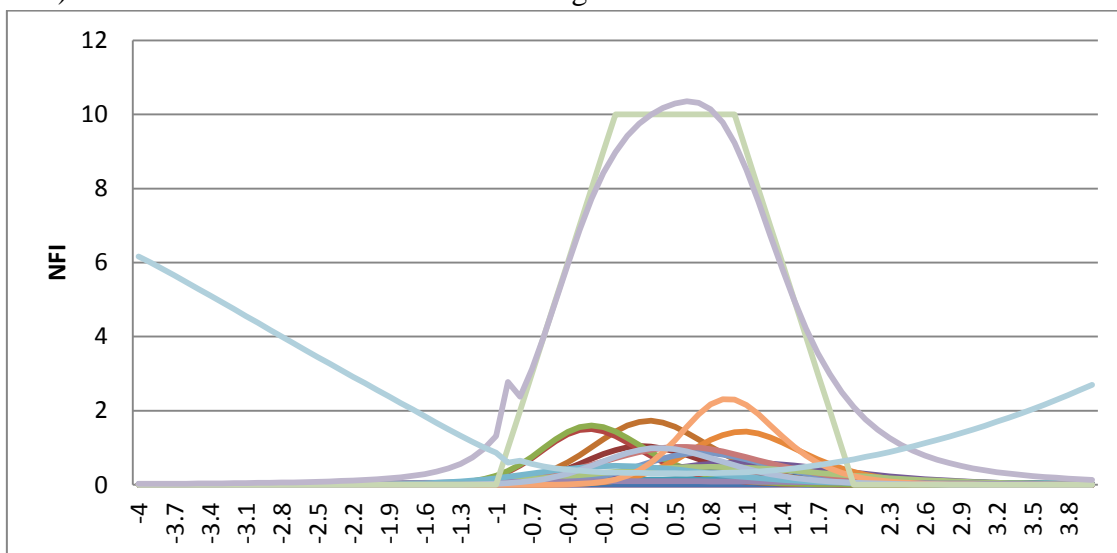

Figure 5

Value ofthe target VIF and test VIF

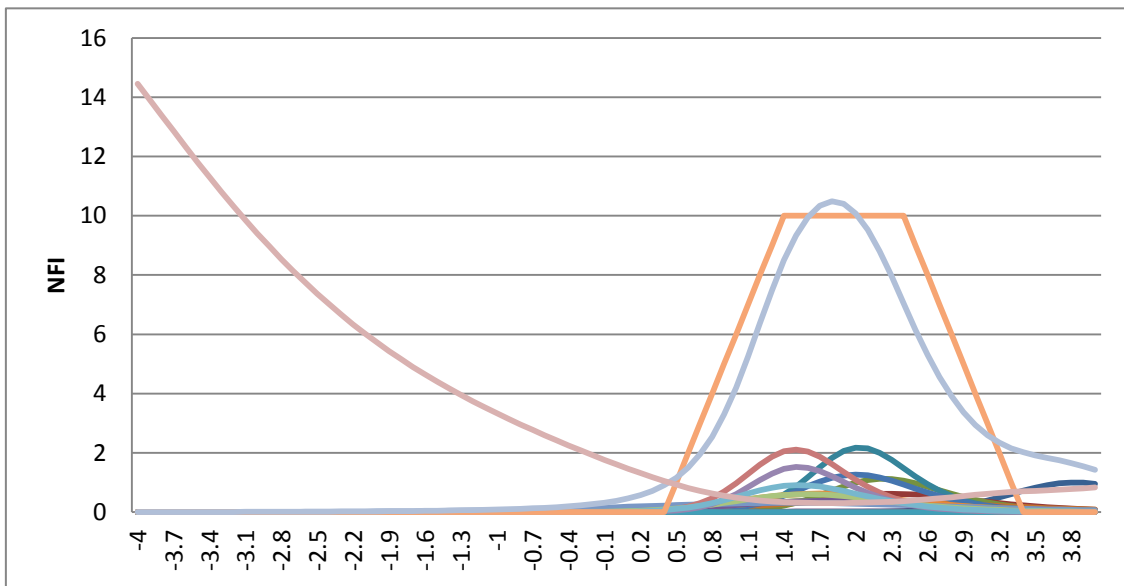and the VIF items that composed for medium ability testee

Figure 6

Value ofthe target VIF and test VIF
and the VIF items that composed for high ability testee

If the VIF target and test are graphed on one figure, it will be resulted the value of the VIF ranged from -4.0 to +4.0 in scale abilities. This suggested that the three test can be used to measure the ability of all students to get good information function value to minimize measurement error. More results are presented in Figure 7.
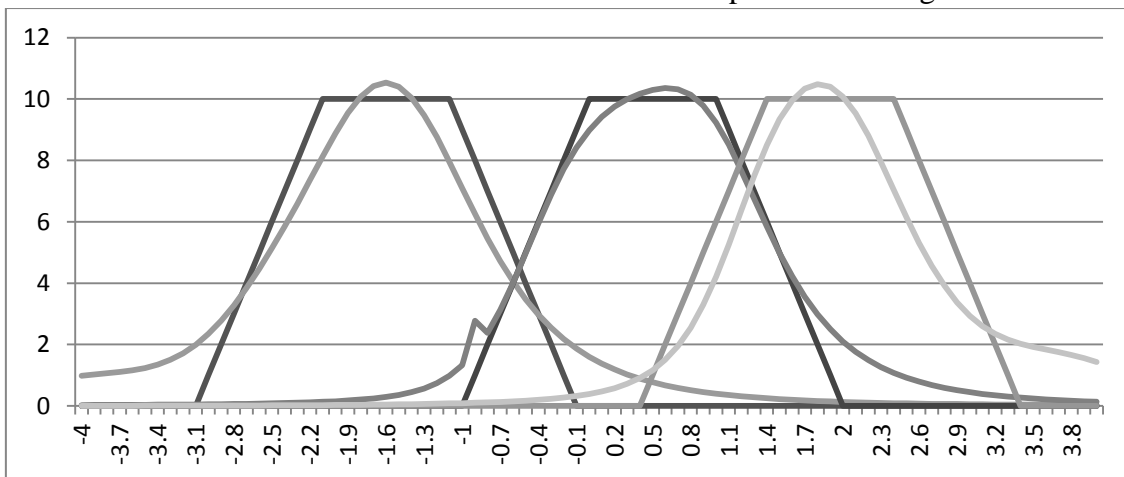

Figure 7

The VIF target and the VIF tests of low, medium,and high ability

Using the item bank contains 200 items which simulated parameters, can be constructed tests the different abilities of testee with the same maximum VIF. For example in this case to the testee are on the low, medium, and high level ability. This test can then be used in accordance with the purpose of test development.

International Seminar on Innovation in Mathematics and Mathematics Education
1st ISIM-MED 2014 Department of Mathematics Education,Yogyakarta State University,Yogyakarta, November 26-30, 2014

EP - 252

Although assembling these tests are based on the VIF and also involved in the item difficulty index in unidimensional item response theory, there are some weaknesses in the process. In assembling the test, a developer remains to consider the content/substance of the item, which can be determined from the indicator of items. So in assembling the test, not merely pursue the achievement of the VIF only, but also it should maintain the content validity of the tests.

By looking at the case of the test assembly at the low, medium, and high level, found that to pursue the VIF target, obtained the different number of the length test. With the difference in the length of test in every level, there have implications for the time of the test. The tests have more items would require a longer time. This needs to be consideration in the test administration.

Conclusion

Assembling the mathematics test from the items have been provided in the item bank requires a special way. One of them is using value of information function (VIF). Each item in the item bank administered its parameters, i.e. the difficulty index, discrimination index, or pseudo guessing. With these parameters, the value of information function can be computed, and based on the value, the standard error of measurement (SEM) can be estimated. The steps of assembling the test using the item bank are determines the desired SEM, specify the target of value of information function, drawing the target of VIF, inserting items from item bank that determine the achievement of the target of VIF, these items are then compiled into a test, using criteria of a good test preparation, for example, sorted from the easiest to the most difficult.

References

Han, K. T. (2007). *WinGen*: Windows software that generates IRT parameters and item responses. Available on www.umass.edu/remp/.../**wingen**/homeF.html. Downloaded in 14 Januari 2014.

Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA : Sage Publication Inc.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA : Kluwer Inc.

Hullin, C. L., et al. (1983). *Item response theory : Application to psichologycal measurement.* Homewood, IL : Dow Jones-Irwin.

Van der Linden, W.J. dan Hambleton, R.K. (1997). Item response theory:brief history, common models and extentions. Dalam Van der Linden, W.J. dan Hambleton, R.K. (Eds). *Handbook of item response theory*. New York : Springer.

Wells, C.S, Hambleton, R.K., Purwono, U. (2009). Test development with IRT models. *Presentation material presented in training IRT collaboration UNY & USAID.*

International Seminar on Innovation in Mathematics and Mathematics Education
1st ISIM-MED 2014 Department of Mathematics Education,Yogyakarta State University,Yogyakarta, November 26-30, 2014

EP - 254