

APPLICATION OF ITEM CHARACTERISTIC CURVE (ICC) IN THE SELECTION OF TEST ITEMS

Adewole Philip (Ph.D.)

Department of Computer Science
University of Lagos. Nigeria

Ojo, Babatunde Odunayo

Senior Assistant Registrar (Research)
The West African Examinations Council
Research Department & WAEC International Office
Plot 6, Lateef Jakande Road, Agidingbi, Ikeja Lagos. Nigeria

ABSTRACT: *Selection of test items is a vital step in test development procedure. Test items could be selected either through the item card or Item Characteristic Curve (ICC). In recent times ICC, has become a vital tool in the selection of tests items as it shows graphically the psychometric properties of the test items. At a glance the curve gives such information as discriminating, difficulty and guessing value (depending on the model). This study works on modeling of 2-parameter model of Item Response Theory (IRT) to generate the item characteristics curve using a software in addition to the windows Microsoft excel application. It also involves a do-it-yourself guide using the windows excel application. The experimental result shows that the ICC curve could be a vital tool in determining the suitability of item for section in a test.*

KEYWORDS: ICC, IRT, 2-parameter model, discrimination index, difficulty index.

INTRODUCTION

A test item could be defined as a question that the testees are expected to respond to. A collection of test items makes a test paper. Test items can be written in various formats: multiple choice, matching, true/false, short answer or essay. Multiple choice, matching and true/false formats are known as Selected-Response Formats (Eleni and Norman, 2005). In these formats, the testees are provided with options from which they choose one. In essay test, the respondents are to give their own responses.

Most public examination bodies use multiple choice questions (MCQ), the essay type, practical tests and in few cases, the test of orals if language testing is involved. The commonest type of test is the MCQ. This may be as a result of numerous advantages that are inherent in it, some of these are that, it could easily be scored, feedback could be instantaneous, the psychometric properties can easily be statistically obtained among others. Kou (2009) confirms that MCQ has scoring rules which do not allow scorers to make subjective inferences or judgments. Every scorer that marks an item, following the rules, assigns the same test score. Each item has options categorised basically into two: the key (correct option) and the distracters (incorrect options). For MCQ, the candidate either gets the full mark for the item or scores zero. The implication is that he/she either

gets it right or wrong. It is unlike the essay question which is subjective to a number of factors. A candidate might get the final answer wrong, but he/she may still get some marks for correct steps that might have led to the correct answer. Scoring, in essay type, requires expertise in the subject area, unlike the MCQ which could be scored without much expertise.

Each test item has its psychometric properties which include test validity and test reliability. Test validity ascertain the test measures what it claims to measure (Paul, 2003). The validity of a test could be determined in a number of ways including content validity, criterion validity and concurrent validity. The reliability of a test, on the other hand, is a consistency or reproducibility of an examinee's performance on the test. A test is reliable, if a student who scores six (6) in a particular test, gets approximately the same score whenever the same test is re-administered on him or her. This is buttress by Colin and Julie (2006), who affirm test reliability as the degree to which an assessment tool (a test) produces stable and consistent results and that test validity refers to how well a test measures what it is purported to measure. They further stated that for a test to be reliable, it needs to be valid. In educational testing, the reliability and validity of a test are key factors to be considered.

Purpose of the Study

Considering the psychometric properties of items before administering it on examinees or selecting it as a test item is a step to having a reliable and valid test items. The application of item characteristic curve (ICC) in test items selection gives a graphical property of the test items. This also gives information about the relationship between the ability of the testees and the probability of a testee getting the item right. The purpose of this study is to present a web application that could help chart the item characteristics curve in addition to giving the interpretation of the curve. In addition to this, a do-it yourself approach of charting the ICC using the Microsoft excel was also explored. The implication of this is that test item moderator will be able to determine the suitability of a test to be selected based on the mature of ICC for the item. It works on modeling of 2-parameter model of Item Response Theory (IRT).

BASIC CONCEPTS IN TEST DEVELOPMENT

(1) Item Analysis

Numerically, MCQ are scored in 1's (ones) or 0's (zeros). The right option is assigned 1 while the wrong options are assigned 0. A candidate's performance in multiple choice test is the ratio of the number of items that he/she has correct response to, to the number of items in the test. In item analysis two statistics (difficulty index and discriminating index) help in analysing the effectiveness of the test items. Difficulty index of an item is the proportion of students who selected the correct options.

Item discrimination or the discriminating power of a test item refers to the degree to which success or failure on the item indicates possession of the ability being measured. It determines the extent to which the given item discriminates among examinees in the function or ability measured by the item. This value ranges between -1 and +1. The higher the value, the more discriminating an item. A highly discriminating item indicates that the students who had high tests scores got the item correct whereas students who had low test scores got the item incorrect (Boopathira and

Chellamani, 2013). Item analysis is a method of reviewing items on a test, both qualitatively and statistically, to ensure that they all meet the minimum quality-control criteria. The main objective is to identify unsuitable items on the test. The analysis is done on each of the items. The value of the analysis has implication on each of the items in the test and not on the test as a whole.

Bharti, Ajita, and Bhavisha (2013) opine that item analysis is the process of collecting, summarizing and using information from students' responses to assess the quality of test items. Difficulty index (b) and discriminating index (a) of each of the items should be known by the subject officer in charge of the compilation of the items for the test. Generally, item analysis shows the following:

- the percentage of the test taking population who got it right;
- the discrimination index of each item: how well item distinguishes between students of high ability and those of low ability;
- the response level of each item: how many actually attempted it, how many got it right and hoe many got it wrong;
- the criterion score on each item: the mean score of all those who did attempt it, and
- whether any distracter did not function well (Kenneth, 2005).

Item analysis investigates the performance of items considered individually either in relation to some external criterial or in relation to the remaining items on the test (Thompson and Levitor, 1985, p. 163). It was further stated that item analysis evaluates the quality of items and of the test as a whole. It can, therefore, be employed to revise and improve on both items and the test as a whole.

(2) **Item Difficulty Index**

In computing item difficulty, items are dichotomously scored (right 1 or wrong 0).

The difficulty index is the proportion of examinees that answer the item correctly:

$$\text{Item difficulty Index} = \frac{\text{number of candidates that got the item right}}{\text{total number of candidates}}$$

The difficulty index ranges from 0 to 1. The higher the difficulty index, the easier the item.

Interpretation

Range	Inference to item	
0.85 – 1.00	Very Easy	
0.70 – 0.84	Easy	
0.30 – 0.69	Optimum	
0.15 – 0.29	Hard	
0.00 – 0.14	Very Hard	Source: Zafar (2008).

Table 1 shows the compilation of difficulty indices for a ten-item test.

Table 1

Difficulty Index for a Ten-Item Test

	Item 01	Item 02	Item 03	Item 04	Item 05	Item 06	Item 07	Item 08	Item 09	Item 10
Candidate 01	1	1	1	1	0	1	1	1	1	1
Candidate 02	1	1	1	1	0	0	1	1	1	1
Candidate 03	0		1	1	0	0	1	1	1	1
Candidate 04	0	0	1	1	0	0	1	1	1	1
Candidate 05	0	0	0	0	0	0	0	0	0	1
Candidate 06	0	0	1	0	0	0	0	0	0	1
Candidate 07	1	0	1	1	1	0	1	0	0	0
Candidate 08	1	1	1	0	0	0	1	0	0	
Candidate 09	0	0	0	0	0	0	0	0	1	1
Candidate 10		0	1	0	0	0	0	0	0	1
Item Score	4	3	8	5	1	1	6	4	5	8
Difficulty Index	0.40	0.30	0.80	0.50	0.10	0.10	0.60	0.40	0.50	0.80

From Table 1, items 3, 7 and 10 with difficulty indices of 0.8, 0.6 and 0.8 are very easy items. Items 5 and 6 with difficulty indices of 0.1 each are difficult items.

(3) **Item Discriminating Index**

$$\text{Discriminating index} = \frac{Up - Lp}{\text{total number of test takers in the upper group}}$$

Up = number of test takers in the upper group that got the item right

Lp = number of test takers in the lower group that got the item right.

In using the Microsoft Excel to compute the discriminating index, we need to sort the candidates' scores either in descending order of the total score or ascending order.

Interpretation

Range	Inference to item
Below 0.19	Poor
0.20 – 0.29	Dubious
0.30 and above	Good

Source: Zafar (2008).

Item discriminating index is a measure that shows how well an item has been able to discriminate between high and low scorers. It is one of the parameters to be considered in selecting test item. It describes to what extent an item distinguishes between the low ability and the high ability test

takers. Its value ranges from -1 to +1. It has a negative value when more low ability test takers get the item right than the high ability test takers. A high value of discriminating index (a) implies that the test is effective. If a =1, it implies that all the candidates in the upper group and none in the lower group answers the item correctly. For a standard item, the minimum acceptable value of a is 0.3 (Zafar 2008).

Table 2 show the discriminating index for a ten-item test for 10 testees.

Table 2

Discriminating Index for a Ten-Item Test

	Item 01	Item 02	Item 03	Item 04	Item 05	Item 06	Item 07	Item 08	Item 09	Item 10	Total	
<i>Candidate 01</i>	1	1	1	1	0	1	1	0	1	1	8	<i>Upper group</i>
<i>Candidate 02</i>	0	1	1	1	0	0	1	0	1	0	5	
<i>Candidate 03</i>	0	0	1	1	0	0	1	1	1	0	5	
<i>Candidate 04</i>	0	0	1	1	0	0	1	1	1	0	5	
<i>Candidate 07</i>	1	0	1	1	1	0	1	0	0	0	5	
<i>Candidate 08</i>	1	1	1	0	0	0	1	0	0	0	4	<i>Lower group</i>
<i>Candidate 06</i>	0	0	1	0	0	1	0	0	0	1	3	
<i>Candidate 09</i>	0	1	0	0	0	0	1	0	0	1	3	
<i>Candidate 10</i>	1	0	0	0	0	0	1	0	0	1	3	
<i>Candidate 05</i>	0	0	0	0	0	0	1	0	0	1	2	
Discrimination	0.00	0.33	1.00	1.00	0.00	0.33	0.00	0.33	1.00	0.67	-	

Note:

- In using Microsoft Excel worksheet, we need to arrange the scores in either descending or ascending order. This would have affected the candidates’ arrangement.
- Upper group comprises of candidates 1, 2, and 3. Lower group comprises of candidates 9, 10 and 5.
- Discrimination index for item 2 = $\frac{2-1}{3} = 0.33$

Table 2 shows that item 10 with a negative value (-0.67) may be wrongly structured, flawed or miskeyed, all (three) candidates in the lower group got the item right while only one candidate in the upper group got it right. Item 9 discriminates very well with a discriminating index of +1. All the candidates (three) in the upper group got the item right while none in the lower group got the item right.

It is expressed on the Item Characteristic Curve as a slope. This is usually denoted by letter (a) in both the 2 and 3-parameter Rasch Models.

(4) Guessing Factor

This is the probability that a candidate that does not know the right answer to the item can guess it right. For a four options item the guessing factor c = (1/4) = 0.25. It is denoted by the value of the lower asymptote on the Item Characteristic Curve.

Table 3

Different Parts of an Item

Question	How many colours has a Rainbow?		
Options	A	7	Key (correct answer)
	B	6	Distracters
	C	5	
	D	4	

Options = suggested plausible answers

Key = Correct answer from the options.

Distracters = other options except the correct answer

Guessing factor (probability) = $\frac{1}{\text{number of options}}$

In the above example, the guessing factor = $\frac{1}{4} = 0.25$

ITEM BANKING

This is an electronic data base for a collection of test items. It is from this collection that selections are made to make up a test. Sarah (2006) defines item banking as a system for collecting and managing collections of assessment items. In his view, Paul (2003) posits that an item bank is a catalogue of test items that are classified according to the content they measure and their psychometric properties.

Items are electronically stored. There is therefore the need for software that can sort and select items based on their psychometric properties. Sarah (2006) posits that a computer can draw items from an item bank based on a set of rules. Annie (1994) asserts that item bank is a collection of test items that may be easily used in compiling test papers. She further states that the use of an item bank and a good item banking software is a solution for the test administrators.

Steps in banking an item

The following are the steps in item banking as elucidated by (Annie, 1994).

1. *Item entry and storage*

This is the process by which items and their psychometric properties and a unique ID for the items are typed in and stored in selected format.

2. *Item retrieval*

This is the electronic selection of items using any of the pre-determined parameters properties or the item ID.

The items previously stored in the item bank have been trial-tested and the difficulty index and the discriminating index have been determined. In order to successfully retrieve items based on psychometric properties earlier discussed, successful Item Calibration is needed.

ITEM CALIBRATION

In selecting items for a particular test, it is expedient to take into consideration psychometric properties like difficulty index, discriminating index and the guessing probability. Item calibration

is the process whereby items are categorised based on these psychometric properties. This is because items are called from the item bank with these properties.

Eggen and Verhelst (2011) posit that efficiency in item calibration is gained when a prior knowledge about the difficulty of the items and the ability of the students are used in allocating students to subsets of items. They quote the works of others about several computer programs with algorithms for item calibration which allow for complete testing designs. For example, BILOG-MG (Zimowski, Muraki, Mislevy, and Bock, 1996), uses the Marginal Maximal Likelihood (MML) approach in the one, two, and three parameter-logistic-test-model and OPLM (Verhelst, Glas and Verstralen, 1995), uses conditional maximum likelihood (CML) as well as MML procedures.

CLASSICAL TEST THEORY

Adedoyin and Mokobi (2013) posit that Classical Test Theory (CTT) is based on the assumption that every individual or person has a true score which is referred to as T and that this true score can only be obtained if the traits are constant and there are no random errors which can affect the result. Andrew (2003) opines that CTT analyses are the easiest and most widely used form of analysis. It is performed on the test as a whole rather than on the individual item.

Table 4 shows the score of 10 candidates in a ten item test.

Table 4

True Scores for a 10-Item-Test for 10 Candidates

	Item 01	Item 02	Item 03	Item 04	Item 05	Item 06	Item 07	Item 08	Item 09	Item 10	Test Score(X)
Candidate 01	1	1	1	1	0	1	1	1	1	1	9
Candidate 02	1	1	1	1	0	0	1	1	1	1	8
Candidate 03	0		1	1	0	0	1	1	1	1	6
Candidate 04	0	0	1	1	0	0	1	1	1	1	6
Candidate 05	0	0	0	0	0	0	0	0	0	1	1
Candidate 06	0	0	1	0	0	0	0	0	0	1	2
Candidate 07	1	0	1	1	1	0	1	0	0	0	5
Candidate 08	1	1	1	0	0	0	1	0	0		4
Candidate 09	0	0	0	0	0	0	0	0	1	1	2
Candidate 10		0	1	0	0	0	0	0	0	1	2

The true score for candidate 01 is 9 that is the summation of the correct items.

Test Score (X) = True score (T) + Measurement Error (E)

In CTT we assume that the Measurement Error is

- normally distributed
- uncorrelated with the true score
- zero

The assumption is that every test score is valid provided the error can be approximated to zero.

ITEM INTERDEPENDENCE (TETRACHORIC CORRELATION)

Items must be checked for interdependency in a bid to making sure that none of the items is serving as a clue to others. Only two items can be checked at once. Tetrachoric correlation is applicable when both observed variables are dichotomously scored. We generate a 2 x 2 matrix for any two items.

$$\text{Tetrachoric correlation (X)} = \text{Cos}\left(\frac{180}{1 + \sqrt{\frac{b*c}{a*d}}}\right)$$

Legend

a= Number of candidates that got both items right

b= Number of candidates that got item first item right and second item wrong

c= Number of candidates that got item first item wrong and second item right

d=Number of candidates that got both items wrong

		Item 4	
		Right	Wrong
Item 5	Right	a	b
	Wrong	c	d

For an illustration, let us consider items 4 and 5 in Table 2.

a=1, b=4, c=0, d=5

Recall

$$\text{Tetrachoric correlation (X)} = \text{Cos}\left(\frac{180}{1 + \sqrt{\frac{b*c}{a*d}}}\right)$$

$$= \text{Cos}\left(\frac{180}{1 + \sqrt{\frac{4*0}{1*5}}}\right) = \text{Cos} 180 = -1$$

Since the Tetrachoric relation coefficient < .49, there is no interdependence between items 4 and 5.

Interpretations

X ≤ .49 Non dependent

X ≥ .50 There exist traces of dependence

Source: Zafar (2008).

ITEM RESPONSE THEORY

The major challenge of the CTT is the measurement error which has influence on the test score of the candidate. Item Response Theory (IRT) was originally developed in order to overcome this challenge with Classical Test Theory (CTT) (Marie, 2004). The CTT focuses on the summative performance in the test (see table 4), while IRT focuses on the individual items. Marie (2004) further posits that one of the basic assumptions of IRT is that the latent ability of a test-taker is independent of the content of a test. Another assumption is that it does not matter which items are used in order to estimate the test-takers' abilities. These assumptions make it possible to compare test-takers' results despite the fact that they have taken different versions of a test (Hambleton and Swaminathan, 1985). In their views, Yen and Fitzpatrick (2006) state that IRT models relate item scores to examinees ability levels and item parameters using non-linear functions. One of the advantages is that the examinees' ability estimates are independent on the particular sample selected for the test (Mustasem and Hassan, 2013). With IRT, items are constructed, administered, and tested for goodness of fit with the model. Items found not fit for the model are eliminated while those that fit the models are easily stored (Ojerinde, Kunmi, Francis and Patrick, 2012).

ITEM CHARACTERISTIC CURVE (ICC)

This is a chart showing the properties of an item in terms of its difficulty index, discriminating index and guessing factor (for a 3-parameter-model). It gives comprehensive information about test item. It proffers solution to some questions like:

- (1) Does the item discriminate well between the good and the poor testees?
- (2) How do the low/high ability testees respond to the item?

Under the item response theory, the standard mathematical model for the item characteristic curve is the cumulative form of the logistic function. The theoretical range of ability is from negative infinity to positive infinity. Practical consideration usually limits the range of values from -3 to +3 (Frank, 2001).

The model after declaration of the parameters can be expressed thus:

$$P(\Theta) = \frac{1}{(1 + \exp(-a(\Theta - b)))}$$

Where Exp (e) is the constant 2.718

b = difficulty parameter

a = discrimination parameter

Θ = ability level (-3 to +3)

Table 5

Types of models and parameters

Model	Parameters used	
1-Parameter	✓	Difficulty index (b)
2-Parameters	✓	Difficulty index (b)
	✓	Discriminating index (a)
3-Parameters	✓	Difficulty index (b)
	✓	Discriminating index (a)
	✓	Guessing factor (c)

Consider an item with a discrimination index of 0.5 and a difficulty index of 0.6. We can use the Microsoft Excel to compute the values of $P(\Theta)$ for different ability levels (-3 to +3).

The equation for the 2-parameter model is

$$P(\Theta) = \frac{1}{(1 + \exp(-a(\Theta - b)))}$$

PRACTICAL DEMONSTRATION USING THE MICROSOFT EXCEL FOR COMPUTATION

Table 6.

Computation of Theta Values for an Item with a= 0.5 and b = 0.6

Thet a	discriminatio n index(a)	difficulty index (b)	L= -a(Θ-b)	exp(-L)	1 +exp(- L)	P= 1/ (1+ exp(L))
-3	0.5	0.6	-1.8	6.050	7.050	0.142
-2	0.5	0.6	-1.3	3.669	4.669	0.214
-1	0.5	0.6	-0.8	2.226	3.226	0.310
0	0.5	0.6	-0.3	1.350	2.350	0.426
1	0.5	0.6	0.2	0.819	1.819	0.550
2	0.5	0.6	0.7	0.497	1.497	0.668
3	0.5	0.6	1.2	0.301	1.301	0.769

Using the Microsoft excel to draw the ICC curve for a 2-parameter IRT model

Open to Microsoft excel

Activity 1

Cell	Type	Cell	Type
A3	Item 1	B1	A
A4	Item 2	C1	B
A5	Item 3		
A6	Item 4		

Activity 2

Cell	Type	Cell	Type
D2	-3	H2	1
E2	-2	I2	2
F2	-1	J2	3
G2	0		

We insert the formula in CELL D3 by first typing equals to sign (=) and
 $= 1 / (1 + \text{EXP}(-\$B3*(D\$2-\$C3)))$

Then place the cursor at the lower end (right) of the cell until a cross (+) sign is shown. Then drag it from D3 to J3.

Repeat the same action for cell D4, D5 and D6

Your output will look like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	a	b	t-3	t-2	t-1	t0	t1	t2	t3					
2				-3	-2	-1	0	1	2	3				
3	Item 1	1	0.9	0.020	0.052	0.130	0.289	0.525	0.750	0.891				
4	Item 2	0.1	0.8	0.406	0.430	0.455	0.480	0.505	0.530	0.555				
5	Item 3	0.5	1	0.119	0.182	0.269	0.378	0.500	0.622	0.731				
6	Item 4	0.5	0.7	0.136	0.206	0.299	0.413	0.537	0.657	0.760				

Figure 1. Microsoft excel output of the generated table for the 2 parameter model (Activity 2).
 Highlight CELL D2 to J6

	a	b	t-3	t-2	t-1	t0	t1	t2	t3
Item 1	1	0.9	0.020	0.052	0.130	0.289	0.525	0.750	0.891
Item 2	0.1	0.8	0.406	0.430	0.455	0.480	0.505	0.530	0.555
Item 3	0.5	1	0.119	0.182	0.269	0.378	0.500	0.622	0.731
Item 4	0.5	0.7	0.136	0.206	0.299	0.413	0.537	0.657	0.760

Figure 2. Highlighted output of the generated table of the 2-parameter model
Select the table as shown above

On the Chart menu- go to -XY Scatter then select smooth Line

(\$ sign tells the computer to pick the value in the cell alone)

$$P(\theta) = 1 / (1 + \text{EXP}(-\$B3*(D\$2-\$C3)))$$

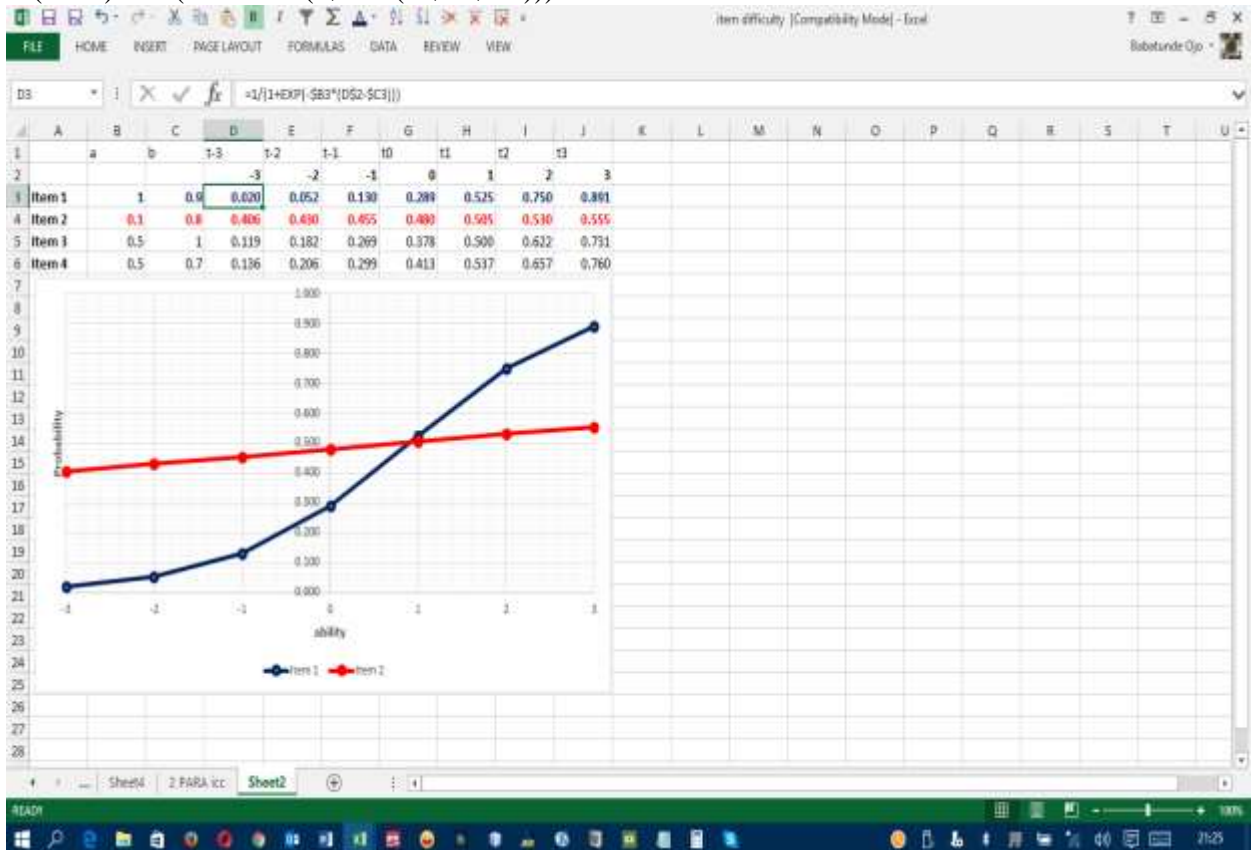


Figure 3: ICC curve for 2-parameters model.

Table 7
 Computed values for a 2-Parameter model for 4-item test

Parameters	Ability								
	Discriminatio n Index	Difficult y index	-3	-2	-1	0	1	2	3
Item 1	1	0.9	0.029	0.052	0.130	0.289	0.525	0.	0.891
Item 2	0.1	0.8	0.406	0.430	0.455	0.480	0.505	0.530	0.555
Item 3	0.5	0.1	0.175	0.259	0.366	0.488	0.611	0.721	0.810
Item 4	0.5	0.7	0.136	0.206	0.299	0.413	0.537	0.657	0.760

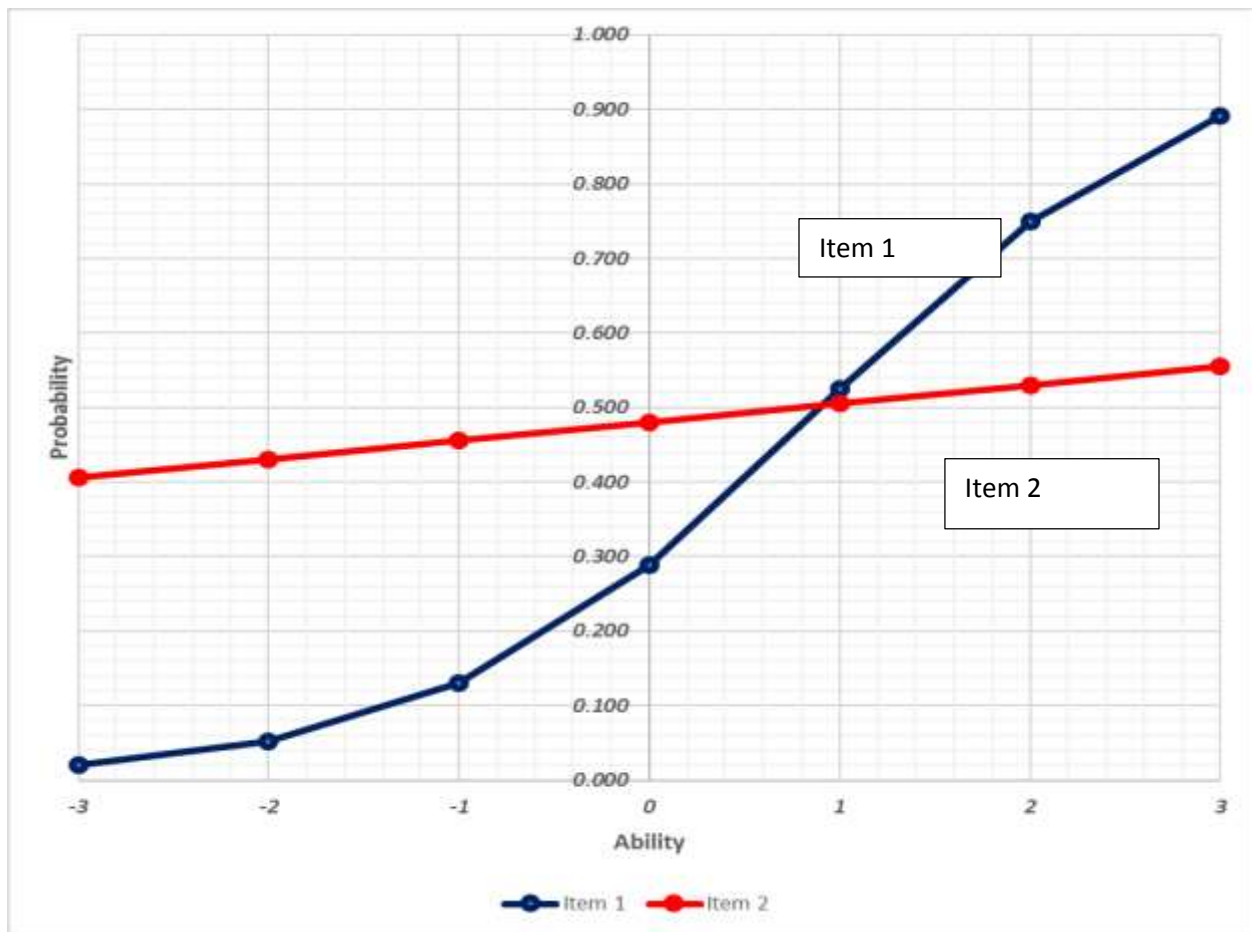


Figure 4: ICC curve for Items 1 and 2 in table 7.

INTERPRETATION OF CURVES IN FIGURE 2

Inference from item 1

Item 1 is much better than item 2.

- (1) Test-takers of low ability (-3) have very low probability of getting the item right.
- (2) It discriminates well between the low and the high ability test-takers.
- (3) It shows that when $Q=0$, the probability of getting the item right is 0.3. The probability of getting the item right increases as the ability (Θ) increases from -3 to 3.

Inference from item 2

Item 2 appears to be a bad item

- (1) Test-takers from the low ability group (-3) have a probability of 0.4 of getting the item right.
- (2) It does not discriminate well between the low ability and the high ability test-takers.

CONSIDER THE FOLLOWING CASE STUDIES

Case 1. Items with the same discriminating index but with different difficulty index.

	Item 1	Item 2	Item 3
Discriminating index (Slope) a	1	1	1
Difficulty index (Location) b	0.1	0.6	0.9

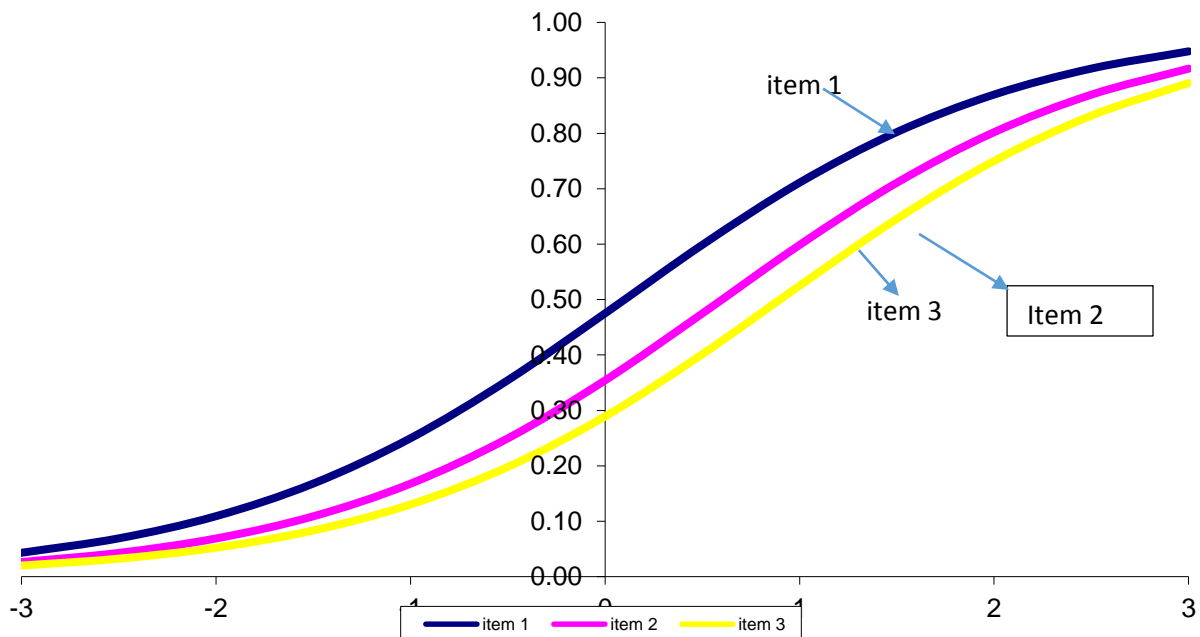


Figure 5. Items with the same discriminating index but with different difficulty index.

Figure 5 shows the ICC for items 1, 2 and 3 with the same discriminating index but different difficulty levels. Item 1 represents an easy item because the possibility of low ability test takers at point zero on the ability scale is highest for the three items. Item 3 represents a difficult item since the probability at point zero on the ability scale is the lowest among the three.

Case 2: Items with the same difficulty index but with different discriminating index.

		Item1	Item2	Item3
Discriminating index (slope)	a	0.1	0.5	0.9
Difficulty index (location)	b	0.8	0.8	0.8

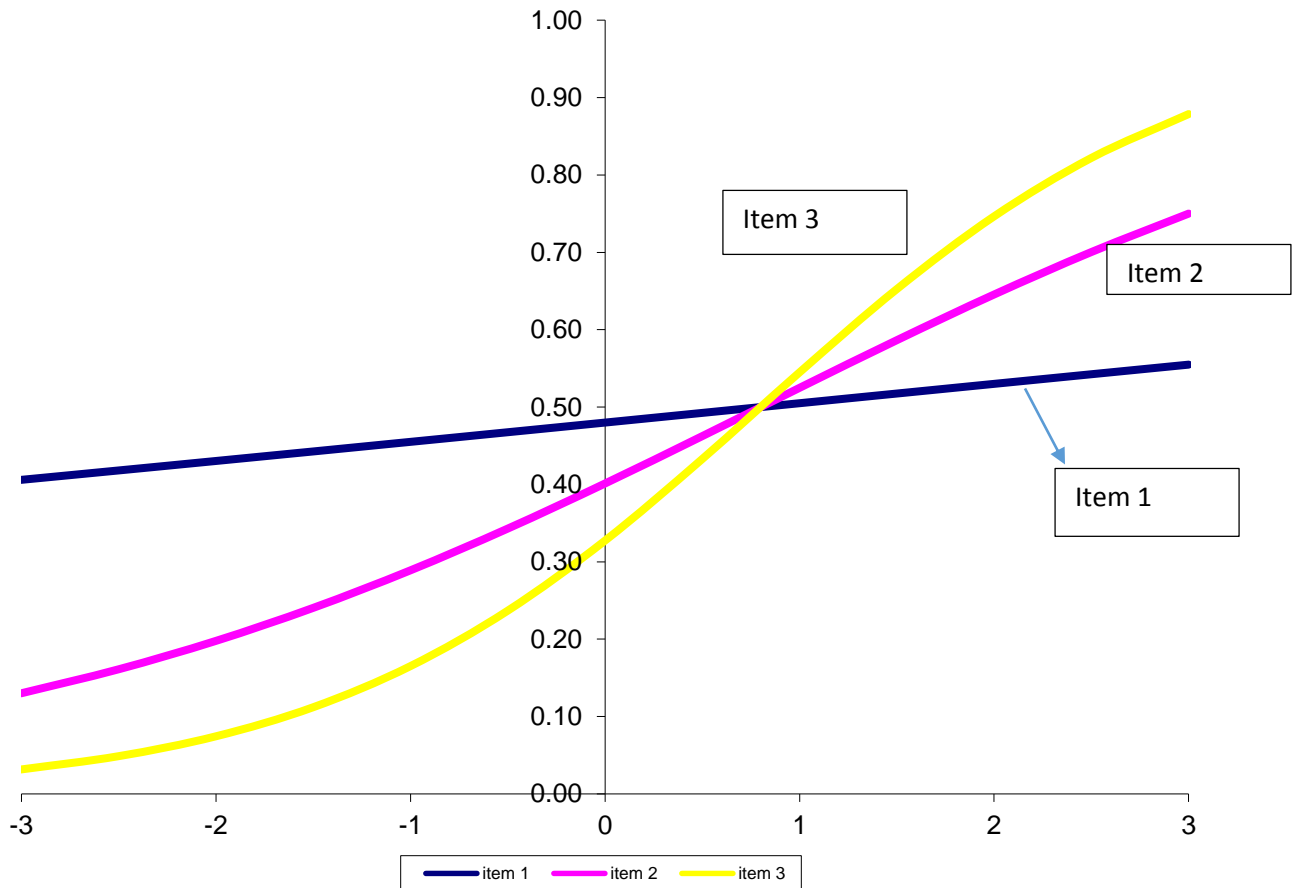


Figure 6. **Items with the same difficulty index but with different discriminating index.**

Figure 6 shows that item 1 is a bad item, as low ability test takers have a high probability of getting the item right. Item 3 appears the best out of the three items since it discriminates very well.

Case 3: Items with negative discriminating index.

		Item1	Item2	Item3
Discriminating index (slope)	a	-0.7	0.4	0.9
Difficulty index (Location)	b	0.8	0.8	0.8

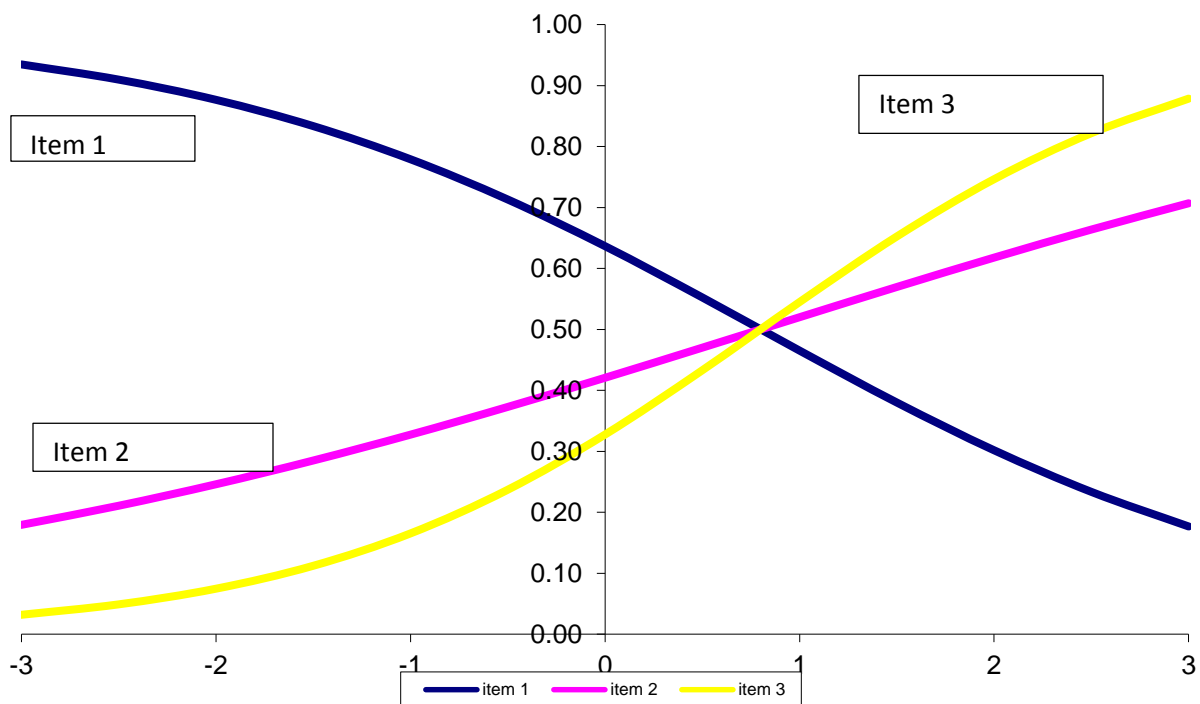


Figure 7. **Items with negative discriminating index.**

Figure 7 shows that item 1 discriminates negatively, low ability test takers have a very high probability of getting the item right. It might have been possibly keyed.

INTERFACE DESIGN AND ACTIVITY

The entire concept of this work has been developed into a specially designed software. For the purpose of this presentation, it has been designed to work on any internet browser on computer system, preferably, Google Chrome browser.

(1) Features of the Interface

- (a) The interface of this design is user friendly.
- (b) It has facility to accept predetermined values of discriminating (a) and difficulty index (b).
- (c) It has facility to accept input for the proposed item and four response options (a,b,c, and d).

- (d) The left hand side accepts predetermined inputs for discriminating index, difficulty index and the proposed item for which the ICC is to be charted.
- (e) The righthand side shows the output of the ICC curve and automatically enlarges on a new window.

(2) ACTIVITY ON DESIGN INTERFACE

- (a) Click on any browser of your choice preferrably Google Chrome.
- (b) In the address bar, type in localhost (in one word)
- (c) Login to the interface.
- (d) The output looks like the figure 9.

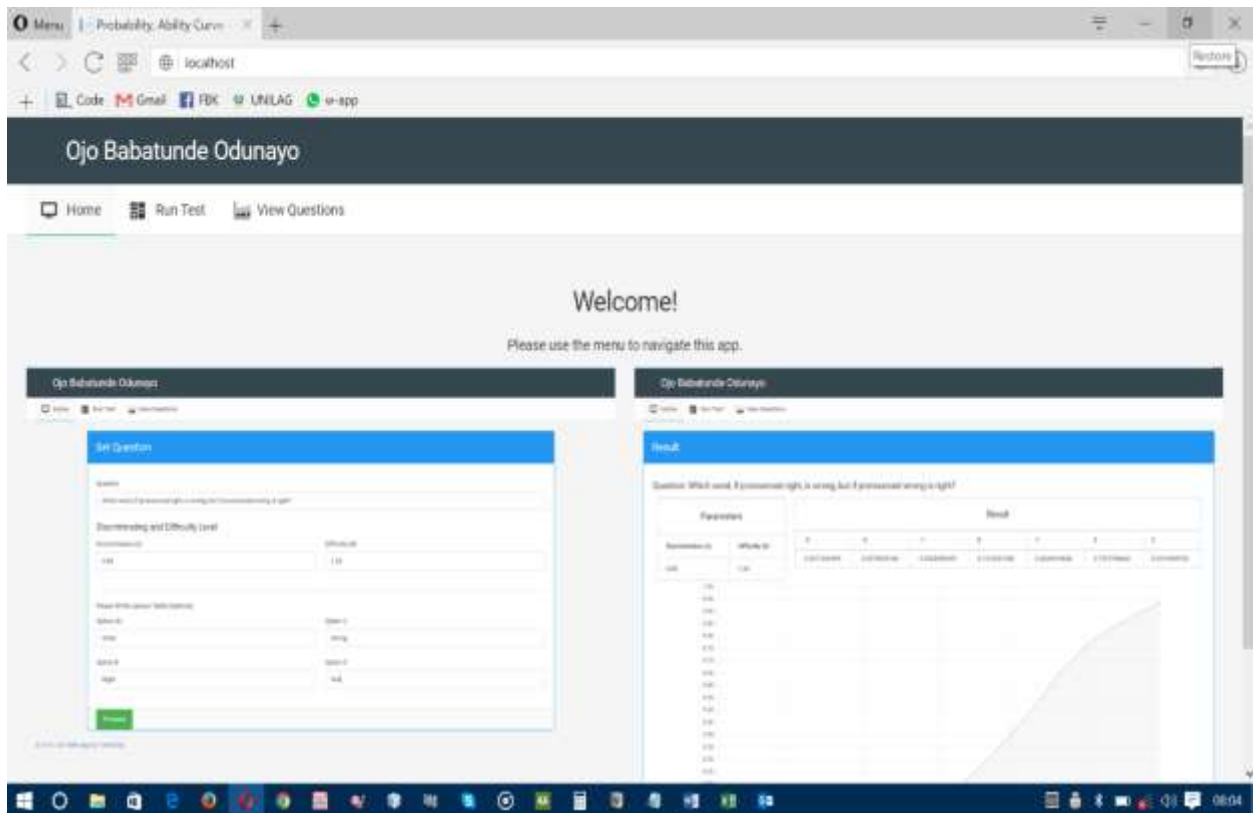


Figure 9. Home page output of Design Interface

- (e) Input the predetermined discriminating (a) and Difficulty Values(b) indices.
- (f) Input the item to be charted using the ICC.
- (g) Input the response options (04) to the question in the spaces provided in the form a,b,c and d.
- (h) Click on Proceed
- (i) The ICC curve is automatically generated on a new window platform with the predetermined parameters a and b.
- (j) The final output of the ICC curve looks like figure 10

Question 1: Lagos State is in which Country?

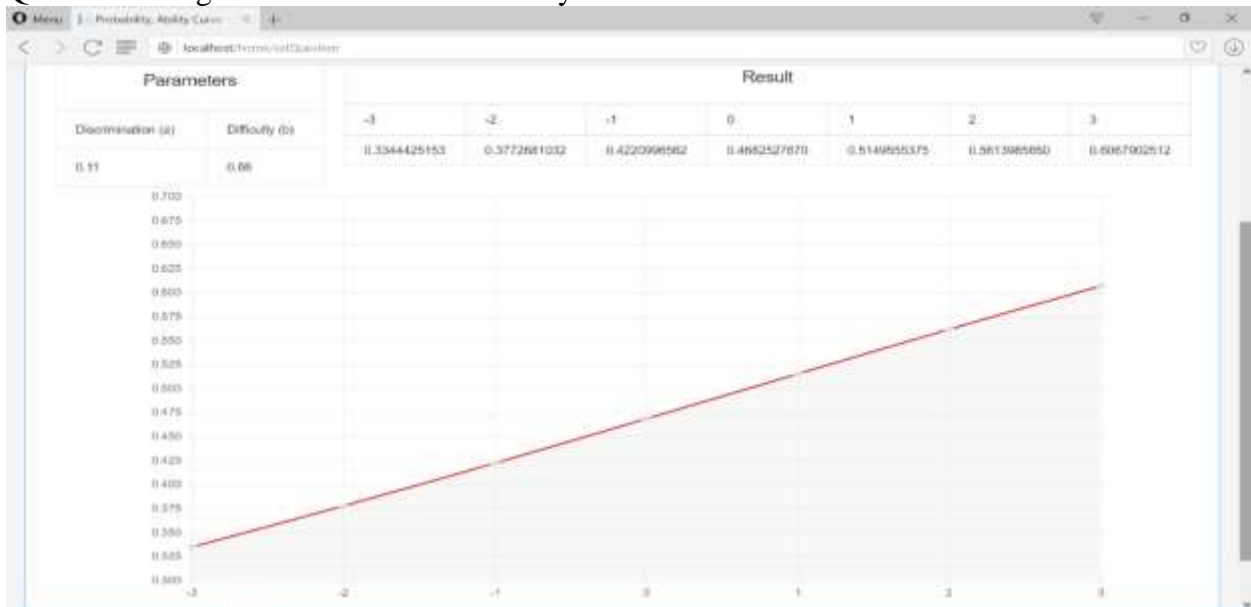


Figure 10. shows the curve generated for the question and parameters inputed.

Question2: If K is a number between 9 and 13, then K + 7 is between what two numbers?

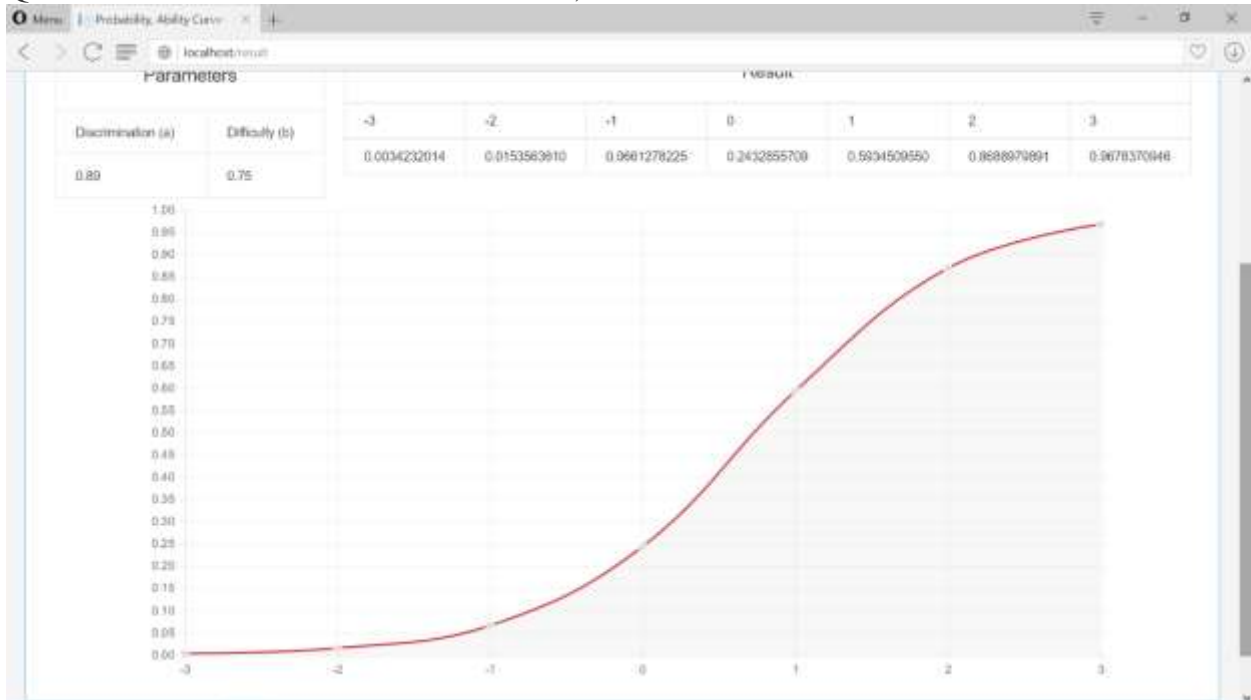


Figure 11. shows the curve generated for the above question. As seen, in fig 11, item 2 is better than item1.

IMPLICATION OF THE STUDY

This study has been able to use a web application to chart the item characteristic Curves (ICC) for items with known difficulty index and discriminating indices. In addition, it has facility like an item bank for all items that have been processed. The study will be of great value to public examination bodies as it gives the psychometric properties of items graphically. This makes decision making on tests to be selected for test takers easy.

APPLICATION OF THE STUDY

The ICC, at a glance, shows the properties of items. It helps the examination body select items that are good, since, at a glance with the nature of the ICC, we can determine the viability of the item to be selected for a particular test paper. The software will generate for each item the ICC using a feedback of difficulty index and discriminating index generated from the IRT scoring model.

LIMITATIONS

This software does not consider Tetrachoric relations (interdependence) between items
This software does not have facility for banking of items.

SUMMARY

This study has been able to demonstrate the computation of the ICC curve of items using the discrimination and difficulty indexes of items. It graphically shows the nature of the ICC, guessing factor (lower asymptote). In addition, previously analysed items could also be reviewed.

RECOMMENDATIONS

- (1) There is the need for public examination bodies to purchase software that could be used to generate ICC.
- (2) There is the need for subject officers of public examination bodies to have ICC for each item to determine the suitability of the items for the intended test.
- (3) There is need for subject officers to be trained on the application of ICC.

SUGGESTION FOR FURTHER STUDY

A further research on this study could be carried out on Tetrachoric relations between two items.

CONCLUSION

It has been successfully shown that the ICC curve generated is a suitable means of determining the choice of items by subject officers in public examination bodies and other related bodies.

REFERENCES

- Adedoyin, O., & Mokobi, T. (2013). Using IRT psychometric analysis in examining the quality of Junior Certificate Mathematics multiple choice examination test items. *International Journal of Asian Science*. Vol 3(4) pp. 992-1011.
- Andrew, A. (2003). *Introduction to item response theory*. Cal State Northridge.
- Annie, W. W. (1994). Guidelines for the development of item banks. International Topics in Educational Measurement (ITEMS). California: Lutheran University and the Thechne Group, Inc.
- Bharti, N. K., Ajita, P., & Bhavisha, N.V. (2013). The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of Year II M.B.B.S students. *International e-Journal of Science, Medicine & Education* Vol. (2): 41-46
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & interdisciplinary Research* Vol. 2 (2).
- Colin, P., & Julie, W. (2006). Exploring reliability in academic assessment. University of Northern Iowa. Retrieved May 23, 2015 from <https://www.uni.edu/assessment/issues.shtml>
- Eggen, T. J. H., & Verhelst, N. D. (2011). Item Calibration in incomplete testing designs. *Psicologica* (2011) 32, pp. 107-132.
- Eleni, D., & Norman, R. (2005, October). Assessment formats: do they make a difference? Centre for Education, University of Glasgow, G 12 8 QQ UK.
- Frank, B. (2001). The basics of item response theory. ERIC Clearinghouse on Assessment and Evaluation.
- Hambleton, R. K., & Swaminathan, H. (1995). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Jason, M., & Judith, A. (1984). Issues in item banking. *Journal of Educational Measurement* 21(4) pp. 315-330.
- Kelly, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30, 17-24.
- Kou, M. (2009). How to construct test items. Retrieved April 7, 2016 from <http://www.catforms.com/pages/Introduction-to-Test-Items.html>
- Kenneth, N.R. (2005). Item writing for tests and examinations. Quantitative research methods on educational planning. UNESCO.
- Marie, W. (2004). Classical test theory vs. Item response theory: An evaluation of the theory test In the Swedish driving-license test. UMEA University.
- Mustasem, A., & Hassan, A. (2013). Empirical investigation of the stability of IRT Item-Parameters estimation. *International Online Journal of Educational Sciences*, 5(2), 291-301
- Ojerinde, O., Kunmi, P., Francis, O., & Patrick, O. (2012). *Introduction to item response theory. parameters models, estimation and application*. Goshen Printmedia Limited
- Paul, S. (2003). Concept paper on an item bank approach to testing. Applied skills & knowledge. Morristown NJ 07960. www.appliedskills.com
- Sarah, C. (2006). Assessment in item banking and repositories. Center for Educational Technology and Interoperability Standards (CETIS) pp. 9-15.

- Thompson, B., & Levitov, J. E. (1985). Using microcomputers to score and evaluate test items. *Collegiate Microcomputer*, 3, 163-168.
- Verhelst, N.D., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). One-parameter logistic model (OPLM). [Computer software]. Arnhem: Cito.
- Yen, W. M., & Fitzpatrick, A. (2006). *Educational measurement* (4th Ed.). American Council on Education and Praeger Publishers.
- Zafar I., (2008). Item analysis assumptions (Difficulty & discrimination indexes). Fact sheet 24A Retrieved July 20, 2016 from https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fcom.ksauhs.edu.sa%2Feng%2Fimages%2FDME_Fact_Sheets%2Ffs_24.doc
- Zimowski, M. F., Muraki, E., Mislevy, R.J., & Bock, R.D. (1996). BILOG-MG [Computer Software]. Chicago: Scientific Software International.