

# International Journal of Information Technology and Business Management

27<sup>th</sup> May 2012. Vol.1 No. 1  
© 2012 JITBM & ARF. All rights reserved



ISSN 2304-0777

[www.jitbm.com](http://www.jitbm.com)

## CONFIDENCE LEVEL RATING IN MULTIPLE CHOICE TEST LEADING TO INCREASE IN RELIABILITY AND DECREASE IN STANDARD ERROR OF MEASUREMENT

**Ms. Niyati Kapoor**

Chitkara International School, Chandigarh, India

**Dr.V.Natarajan,**

Emeritus, Del, India

Email: [niyati@chitkara.edu.in](mailto:niyati@chitkara.edu.in), [drvnatarajan@merittrac.com](mailto:drvnatarajan@merittrac.com)

### ABSTRACT

In the present scenario, whenever any test taker is measured, evaluated and assessed, the number of right answers that a test taker provides, forms the basis of his assessment without understanding the fact that for responding a particular test item how much sure s/he has been. Also to be sure about the knowledge of the test taker, it is important to understand whether the test taker is confident about answering it or has the correct answer been the result of mere guesswork. This Classical Test Theory, applied Test and Item Analysis has a very low reliability and high standard error. A new and innovative technique can be introduced in the normal testing scenario which is called the Confidence Level Rating (CLR) which indicates the confidence s/he has in her/his response. Thus the purpose of this paper is to present how beneficial the confidence level rating is in the multiple choice test items and how this technique leads to increase in reliability and thereby reducing the standard error of measurement. Also this research paper is a by-product of an illuminative research project done by the first author and guided by the second author which offers various innovative practices in the field of learning and assessment which are conducive to the learning process of a learner in school education or otherwise ,one of it being Confidence Level Rating.

### KEY WORDS

Confidence level rating, Confidence-based learning, assessment of learning, assessment for learning, number right scores.....

### 1. INTRODUCTION

In Confidence Level Rating, a test taker indicates the confidence he has in his response. Confidence Level Rating can be used in the assessment of learning as well as in the

assessment for learning. Confidence Level Rating when used in assessment for learning is called Confidence-Based learning. Before understanding

CLR, it is important to understand the two concepts with their meanings. These are:

**1.1 Assessment of Learning**-The 'assessment of learning' is defined as a process whereby someone attempts to describe and quantify the knowledge, attitudes or skills possessed by another. Teacher direction is paramount and the student has little involvement in the design or implementation of the assessment process in these circumstances.

- 1.1.1 Learning is designed by the teacher
- 1.1.2 Evidence is collected by the teacher
- 1.1.3 Whatever has been learnt (and whatever has not been learnt) is judged by the teacher

**1.2 Assessment for Learning**-The 'assessment for learning' involves increased level of student autonomy, but not without teacher guidance and collaboration. The assessment for learning is sometimes seen as being akin to 'formative assessment'. More emphasis is laid towards giving useful advice (feedback) to the student and less emphasis is laid on the giving of marks and the grading function. Whatever has been learnt is judged by the teacher (insight to whatever has not been learnt is developed by the student). AFL is both diagnostic and formative bases on a variety of information sources. Both written and verbal feedback is descriptive. Lesson plans are modified by the teachers according to the feedback. There are no scores or grades for AFL. Record keeping is descriptive. Assessment for learning is continuous as it occurs throughout the learning process in every stage. Assessment for learning works on its five principles mentioned below:

- effective feedback of students
- active involvement of students in their own learning
- adjusting teaching to take account of the results of assessment
- recognition of the profound influence assessment has on the motivation and self-esteem of the pupil
- need for the students to be able to assess themselves and understand how to improve

Now, going back to Confidence-Based learning, it is a methodology used in learning that

measures a learner's knowledge quality by determining both the correctness of the learner's knowledge and confidence he has in that knowledge. The Confidence-Based Learning Methodology is a result of more than 70 years of academic, commercial, and governmental research into the connection between confidence, correctness, retention, and learning. The first academic paper on the subject was written in 1932 and asserted that measuring confidence and knowledge was a better predictor of performance than measuring knowledge alone, which can be prone to guesswork.

When Confidence Level Rating is used in assessment of learning, then it helps in assessing learner's overall performance by judging him according to correctness of the answer as well as the confidence of the test taker in his answer. It is better than assessing just his ability.

In one of the researches, it was found that the questions which were answered correctly were having confidence ratings in the range 40% to 81%. It was also found that there was an association between **increasing overconfidence** and **lower examination scores**. So, it can be concluded that the test takers who are over confident about their knowledge tend to give wrong response.

It is also found in one of the researches that a quicker response to an item is more likely to be correct and a quicker response also indicates a higher confidence rating.

As we know that when assessment of a student/learner is done after a specific period of learning to know if he had acquired the knowledge and intellectual abilities in that period of time or not, then it is called assessment of learning.

Now, in the present situation, when assessment of a student is done in the form of a multiple choice test, he is supposed to choose an option for a particular item out of the four or five options given to him. If he chooses the correct option, he gets 1 mark but if he chooses the wrong option, he gets 0 (or sometimes negative) marks. In this scenario, it can't be found out if the person is guessing or not i.e. it can't be found out if he is sure of his knowledge or not. And in case, he is guessing, what is the extent to which he is guessing. All these factors remain unidentified in

a normal multiple choice test. The formula scoring (score corrected for guessing = no. right score – the no. of wrong choices/ n-1 where n = the no. of options in every item). The formula score is more valid when blind guessing occurs. There can be two other ways of guessing namely, Informed Guessing (less information) and Intelligent Guessing. These two are not accounted for a negative marking. Also, an assumption in the formula score is that the wrong responses are due to guessing. This presupposes that the correction for the same is taken as penalizing the wrong answers as if all wrong answers are because of guessing. Guessing may be due to ignorance and it is not the same as punishing for wrong answers thinking that all wrong answers are due to guessing.

To know the guessing factor or to understand the extent of guessing for a multiple choice item, **confidence level rating** can be introduced. In confidence level rating, the test taker indicates how sure he is that his response is correct on a four or five point scale. Confidence of a test taker in his knowledge basically enables us to understand the difference between what he thinks he knows and what he actually knows. In the world, many have adopted the confidence level rating to understand how much a test taker is sure of his response and different scales are developed to rate their confidence level. Some of these are as follows:

- 0%, 25%, 50%, 75% and 100% (5 point scale).
- Guess, low confidence, medium confidence and high confidence (4 point scale).
- Certain, confident and unsure (3 point scale).
- Correct, wrong and not sure (3 point scale).
- Absolute Confidence (AC), Partial Confidence (PC) and Random Guessing (RG) (3 point scale).

However, the following scale of different ranges can be developed to rate the confidence level of test takers:

## 2. OBJECTIVES

- To investigate into the basic concepts of Confidence Level Rating in learning and testing situations. (Literature Survey).

- 0% to 25%.
- 26% to 50%.
- 51% to 75%.
- 76% to 100%.

The test taker indicates his/her surety for each of the items of the test along with the answer he chooses.

So, marks can be allotted to test takers accordingly by taking in account both: **the correctness of the answer** and **his surety (confidence level)** about his knowledge.

The use of confidence rating also indicate the extent to which the students are over-estimating or under-estimating their skills.

Confidence level rating is also very helpful when **recruitment tests** are concerned. In any job, the confidence with which a candidate performs at workplace is equally important to the knowledge and ability he needs to perform the job efficiently. The following cases could be considered:

- Those who have **high confidence** and **less knowledge** are more likely to make mistakes on the job and put the companies on risk.
- Those who have **low confidence** and **more knowledge** are knowledgeable but are not sure of their knowledge and thus they are hesitant in their actions. They are likely to act with hesitation or don't act at all on the job.
- Those with **low confidence** and **low knowledge** will not at all perform productively or effectively on the job.
- Those with **high confidence** and **high knowledge** are likely to act and act correctly resulting in higher and productive performance.

- To explore the different ways of conducting Confidence Level Rating.

- To find out the best practice of Confidence Level Rating in the world so far.
- To come out with a method to apply the concept of Confidence Level Rating practically in learning and assessment situations, and to illustrate it with the help of an example.
- To understand the relevance of Confidence level Rating at primary level 4<sup>th</sup> 5<sup>th</sup> and 6<sup>th</sup> graders in school education.
- To provide an appropriate feedback to the test takers with the adoption of Confidence level rating.

### 3. METHODOLOGY

This paper proceeds with the process of defining the confidence level of a test taker along with providing the reasons to suggest the need to apply CLR in learning and testing along with assessment of abilities. This also proceeds clarifying the gaps in the present assessment system of education thereby suggesting an innovative technique to make the assessment learner friendly and productive. Also a different thought process gets incorporated while applying the CLR where the understanding of CLR in the areas of learning as well as assessment for learning becomes imperative and hence the analysis of the same is done accordingly to make it useful and applicable. The next step is the development of appropriate tool and scale for measuring confidence level of test takers and also designing of a new method of measuring confidence of test takers along with the assessment of their knowledge relating to online examinations. After the study of CLR for assessment of learning the next step is illustration

of its outcomes. This study also adds knowledge related to the guess work done by the test taker while attempting a particular test item hence throwing light on the gap between one's knowledge and the surety with which he attempts a test item relating to his knowledge about a particular concept.

Further, this proceeds with impact of CLR on the feedback provided to the test taker post a test, whereby affecting his psychology altogether. Ultimately the study examines the effect of feedback given to a student according to his confidence in retention of his knowledge. Finally it deals with the prediction of the future performance of candidates appearing for recruitment tests for any job according to their Confidence Level Rating and also has a potential to recommend the assessment companies to use CLR along with assessment for better assessments.

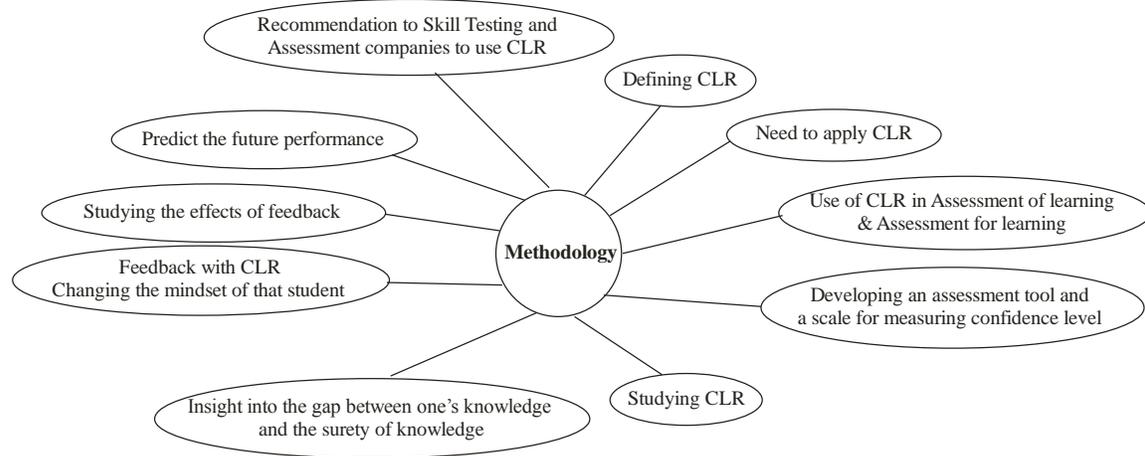


Figure 3.1: Methodology involved in Confidence Level Rating

### 4. THE EXPERIMENT

A Science test was conducted at the Chitkara International School, Chandigarh, India, for Grades 4, 5 and 6 by Ms. Sneha\* and other team

members from MeritTrac (India's largest skill assessment company that provides Testing and Assessment services for corporate sector,

educational Institutions, Government & Public sector). This test was conducted in Grade 4A and 4B of Grade 4, Grade 5A and 5B of Grade 5 and in Grade 6. The test was containing 30 MC items with 4 options in each item in each grade. Same question paper was used for Grade 4A and 4B. Similarly, same question paper was used for Grade 5A and 5B.

Several marking schemes were designed to mark a test taker on his/her correct and incorrect responses in accordance with the Confidence Level Rating he/she provides for all the items in the test. MeritTrac has come up with the following marking scheme for correct and incorrect responses for all the items:

Table 4.1: Showing Rating Scale and Criteria of Marking for Confidence Level Rating

Rating Scale	Marks	
4-point scale:	<b>For Correct Answer</b>	<b>For Incorrect Answer</b>
1) 0% to 25%	1) 0	1) 0
2) 26% to 50%	2) 1	2) -1
3) 51% to 75%	3) 2	3) -1.5
4) 76% to 100%	4) 3	4) -2

After the discussion amongst all the marking schemes, it was clear that this marking scheme will be better and after analyzing the results of tests given in Grade 4, 5 and 6 of Chitkara International School, it confirms that it is a better marking scheme of all the marking schemes.

Here, for the confidence rating of 0% to 25%, test taker is given 0 marks, no matter if the response is correct or incorrect because it seems that the test taker is guessing the answer as he has no confidence in his answer.

#### 4.1 For correct response:

For the confidence rating of 26% to 50%, the test taker is given 1 mark/credit point, for confidence rating of 51% to 75%, he is given 2 marks/credit points and for rating of 76% to 100%, 3 marks/credit points.

Here the marks/credit point's increases as the confidence increases for the correct answer as

credit should be given to confidence level of the test taker.

#### 4.2 For incorrect response:

For the confidence rating of 26% to 50%, the test taker is given -1 mark/credit point, for confidence rating of 51% to 75%, he is given -1.5 marks/credit points and for rating of 76% to 100%, -2 marks/credit points.

Here negative marks/credit points are allotted for the wrong responses so that a test taker is true in providing his/her confidence level. If negative marking is not included then the test taker may think that whatever may be the response, if 76% to 100% rating is given, then marks will be more and so he won't be true in giving the confidence rating.

So, the more is the confidence rating for wrong response, the more negative are the marks.

## 5. FINDINGS

Findings from test and item analysis of CLR statistics of MC test given to primary school children of 4<sup>th</sup>, 5<sup>th</sup> and 6<sup>th</sup> grade in the subject area of Science at Chitkara International School are as follows:

The procedure involved in administering a 30 item test to grades 4, 5 and 6 midterm and studying the responses in two kinds of format: one without CLR and the other with CLR means

that the 2 matrices are generated with (1,0,X), 1 being for the correct response, 0 for incorrect response and X for omitted response. The no. right scores, the odd numbered item scores, even numbered item scores are separately obtained to deal with. The total no. right scores are passed through descriptive statistics obtaining mean, median, mode, SD, variance and other related statistics normally associated with descriptive. Then CLR is looked at for every item for every

test taker and depending on the measurement framework of incorporating CLR separately for

wrong answer and correct answer is as follows:

*Table 5.1: Showing measurement framework of incorporating CLR for wrong answer and correct answer that has been found to be the most effective marking scheme amongst many*

CLR Rating Scale	Marks/Credit points	
4-point scale:	<b>For Correct Answer</b>	<b>For Incorrect Answer</b>
1) 0% to 25%	1) 0	1) 0
2) 26% to 50%	2) 1	2) -1
3) 51% to 75%	3) 2	3) -1.5
4) 76% to 100%	4) 3	4) -2

Accordingly, the CLR scores are obtained for both, no. right scores and odd and even numbered test scores. With the help of this, further

Descriptive Statistics is worked out and the Descriptive Statistics of both are given below:

**Grade 4**

*Table 5.2: Descriptive Statistics of no. right scores*

<b>Descriptive Statistics of no. right scores</b>	
Mean	17.34883721
Standard Error	0.732026209
Median	17
Mode	21
Standard Deviation	4.800216865
Sample Variance	23.04208195
Kurtosis	-0.262465335
Skewness	0.493607012
Range	21
Minimum	8
Maximum	29
Sum	746
Count	43
Largest(1)	29
Smallest(1)	8
Confidence Level (95.0%)	1.477288681

*Table 5.3: Descriptive Statistics for CLR*

<b>Descriptive Statistics for CLR</b>	
Mean	26.06976744
Standard Error	3.208799734
Median	24.5
Mode	32
Standard Deviation	21.04150699
Sample Variance	442.7450166
Kurtosis	-0.027847939
Skewness	0.457643409
Range	96.5
Minimum	-15
Maximum	81.5
Sum	1121
Count	43
Largest(1)	81.5
Smallest(1)	-15
Confidence Level (95.0%)	6.475619955

**Grade 5**

*Table 5.4: Descriptive Statistics of no. right scores*

Mean	14.08
Standard Error	0.761402215
Median	14
Mode	18
Standard Deviation	3.807011076
Sample Variance	14.49333333
Kurtosis	-0.825567504
Skewness	0.277852538
Range	14
Minimum	8
Maximum	22
Sum	352
Count	25
Largest(1)	22
Smallest(1)	8
Confidence Level (95.0%)	1.571456926

*Table 5.5: Descriptive Statistics for CLR*

Mean	10.66
Standard Error	3.355558
Median	8.5
Mode	-0.5
Standard Deviation	16.77779
Sample Variance	281.4942
Kurtosis	-0.18578
Skewness	0.46814
Range	69.5
Minimum	-20
Maximum	49.5
Sum	266.5
Count	25
Largest(1)	49.5
Smallest(1)	-20
Confidence Level(95.0%)	6.92553

**Grade 6**

*Table 5.6: Descriptive Statistics of no. right scores*

Mean	16.09090909
Standard Error	0.717466354
Median	15
Mode	14
Standard Deviation	3.365215495
Sample Variance	11.32467532
Kurtosis	-0.233878644
Skewness	0.746506715
Range	12
Minimum	12
Maximum	24
Sum	354
Count	22
Largest(1)	24
Smallest(1)	12
Confidence Level(95.0%)	1.492052958

*Table 5.7: Descriptive Statistics for CLR*

Mean	19.63636364
Standard Error	3.216600242
Median	12.5
Mode	10
Standard Deviation	15.08719247
Sample Variance	227.6233766
Kurtosis	-0.42569633
Skewness	0.798270913
Range	53.5
Minimum	-1
Maximum	52.5
Sum	432
Count	22
Largest(1)	52.5
Smallest(1)	-1
Confidence Level (95.0%)	6.689286372

Continuing the analysis, reliability estimates are done through KR21, KR20 and split half – full test reliability.

KR21 reliability is the lowest bound estimate while split half, KR20 are the middle values and

the highest bound estimate is the Rulon estimate which is not calculated. The corresponding standard errors of the measurement are then calculated.

Finally, the following table shows the comparison of all the values of test analysis of

Grade 4, 5 and 6 including CLR and excluding CLR:

Table 5.8: Showing comparison of all the values of test analysis of Grade 4, 5 and 6 including CLR and excluding CLR

Item		Grade 4	CLR 4	Grade 5	CLR 5	Grade 6	CLR 6
<b>Mean</b>		17.34883721	26.06976744	14.08	10.66	16.09090909	19.63636364
<b>Median</b>		17	24.5	14	8.5	15	12.5
<b>Standard Deviation</b>		4.800216865	21.04150699	3.807011	16.77779	3.365215495	15.08719247
<b>Variance</b>		23.04208195	442.7450166	14.49333	281.4942	11.32467532	227.6233766
<b>No.Rt. Score</b>	<b>Minimum</b>	8	-15	8	-20	12	-1
	<b>Maximum</b>	29	81.5	22	49.5	24	52.5
<b>Range</b>		21	96.5	14	69.5	12	53.5
<b>T Score</b>	<b>Minimum</b>	30.52413574		34.02947		37.72903741	
	<b>Maximum</b>	74.27215919		70.80372		72.86861211	
<b>T Mod Score</b>	<b>Minimum</b>	18.83861718		24.44715		30.36645985	
	<b>Maximum</b>	88.83545471		83.28596		86.58977937	
<b>Scoring Wt. Scores</b>	<b>Minimum</b>	10.45404208		11.04		14.54545454	
	<b>Maximum</b>	40.22978959		31.44667		32	
	<b>Mean</b>	23.13310927		19.08053		20.71428571	
	<b>Standard Deviation</b>	6.805221434		5.484387		4.654024803	
<b>T Score</b>	<b>Mean</b>	50		50		50	
	<b>Standard Deviation</b>	10		10		10	
<b>T Mod Score</b>	<b>Mean</b>	50		50		50	
	<b>Standard Deviation</b>	16		16		16	
<b>Reliability</b>	<b>KR21</b>	0.706023808	0.971877641	0.501173	0.981188	0.373377157	0.957750702
	<b>SEM</b>	2.60265701	3.528602305	2.688804	2.301186	2.703260239	3.101117194
	<b>KR21(200)</b>	0.941214186	0.995678341	0.870097	0.997132	0.798888869	0.99342654
	<b>SEM(200)</b>	1.163850311	1.383254455	1.372128	0.898458	1.531449919	1.223222464
	<b>KR20</b>	0.756182049	0.998247434	0.636967	0.991229	0.507377576	0.986002452
	<b>SEM</b>	2.370247502	0.880874597	2.29381	1.571308	2.396855397	1.784984379
	<b>KR20 (200)</b>	0.953866321	0.999736723	0.921242	0.998674	0.872876212	0.997875086
	<b>SEM(200)</b>	1.031026678	0.341415659	1.068394	0.610846	1.217581828	0.695471187

A keen observation of the above table shows how the reliability values are increasing when are the method of CLR is included in the test. For the tests including CLR, the values of KR21 are coming to be around 0.99 for every grade which shows how effective is this method of including CLR in any test. It may be observed that there appears to be no need to increase the no. of items from 30 to 200 to check for ETS standard of reliability naming 0.94 and above. Since the CLR reliabilities are much more than 0.94.

It should be reiterated here that children of 4<sup>th</sup>, 5<sup>th</sup> and 6<sup>th</sup> grades have recorded the confidence level very honestly and dedicatedly. It should be noted that any confidence level rating given by ant test taker to any item must be a correct reflection of the confidence of the test taker. Elsewhere in another piece of research, one of the authors of this research is to explore whether this phenomenon is observed with adolescent and adult test takers. The impact of this research is to be felt in preprimary, primary and secondary.

## 6. DISCUSSION (APPLICATION WITH FEEDBACK)

When Confidence Level Rating is used in the assessment for learning, then it helps in giving a better feedback to the students which helps them in planning their learning process and also let them know how cautiously they should read before answering any question. They can also be told at the end of the test if they are over confident or under confident about their learning.

In fact, they can also be given feedback for each item in between the test so that they can know if they need to concentrate more before answering or not. In a computer based test (CBT), after the test taker has selected his response, another window can pop-up asking for their confidence level for that particular answer. And then, when their confidence level is provided, they can be given a feedback regarding their over confidence or under confidence.

In one of the researches, it is found out that by giving the feedback to the students about their knowledge and how confident they are about their knowledge, the retention of knowledge in those students' increases.

Suppose, at the end of the test, a student is given a feedback like he is over confident about his knowledge, then his mindset will change and he will learn more attentively and will answer next time more cautiously. So, in this way, knowing the student's confidence level along with their knowledge and then giving them the feedback will help students a lot for planning their learning process in the future.

In a research done on how Confidence Level Rating helps in assessment for learning, it was found that feedback not only increases the retention of low confidence correct responses, but also improves the accuracy of confidence judgments. And it was also found that feedback helps learners correct memory errors.

The above research findings will help to provide realistic feedback to a test taker. The organizations to which this research finding that will help to first of all innovate an assessment framework for schools involved with i Discoveri and Wipro Foundation.

## 7. REFERENCES

1. Butle, A. C., Karpicke, J. e., & Roediger, H. L. (2008). Correcting a Metacognitive Error. *Journal of Experiment Psychology: Learning memory & cognition*, Vol. 34, No. 4, 918-928.
2. Dr.V.Natarajan, Madan, M., & Mohan, S. (2011). *A Report of Research and Investigation into Confidence Level Rating along with item responses and its impact on scores of test taker*. Delhi: skill assessment company.
3. Hausman, C. L., Weiss, J. C., Lawrence, J. S., & Zeleznik, C. (1990). Confidence weighted answer technique in a group of pediatric residents. *informa health care*, 163-168.
4. HenrikHaahr, J., & Hansen, M. E. (2006). Adult Skills Assessment in Europe Feasibility Study. *Danish Technological Institute*, 1 – 127.
5. Hojat, M. r., Zeleznik, C., Goepf, C. E., Borenstein, M. a., Amadio, P., & Kowlessar, O. (1988). Student's Certainty during Course Test taking and Performance on Clerkships and Board Exams. *academic medicine*.
6. KemiAdeboye, & FintanCulwin. (2003). Validation of The Effectiveness of A Confidence Based Multiple Choice Examination System. *School of Computing London South Bank University*, 44-48.
7. Nickerson, R. S., & McGoldrick, J. C. (1964). Confidence Ratings And Level Of Performance On A Judgmental Task. *Electronic Systems Div L G Hanscom Field Ma Decision Sciences Lab*, 311- 316.
8. Omirin, M. (2007). Validity and Reliability indices of Three Multiple Choice Tests Using the Confidence Scoring Procedure. *medwell journals*, 20-23.
9. Schraw, G. (1997). The Effect of Generalized Metacognitive Knowledge on Test Performance and Confidence Judgments. *The University of Nebraska-Lincoln*, Vol. 65.

10. Shizuka, T. (n.d.). *xploring Time Related Variables In a Computer Based Reading Test*. Japan: Kansai University.
11. Sie Hoe, L., Kian Sam, H., NgeeKiong, L., & Hasbee, U. (2008). Web based Computer-Adaptive Multiple Choice Assessment (CAAS) Using the New NRET Scoring Method. *Asian Technology Conference in Mathematics*. Bangkok: atcm.