A Generic Framework of Computer Adaptive Testing Mechanism

Kiran Pala

International Institute of Information Technology Hyderabad,

Hyderabad, India.

kiranap@gmail.com

Abstract: *Assessment and testing is directly related to education and training. It is a way to measure the level of performance and to understand the comprehension of the students in a group. In this era, using computer technology in testing and assessment has become one of the most important and considerable factor. While constructing a mechanism of Computer Adaptive Testing (CAT), different parameters should be taken into account. In this paper we have focused on a generic framework of a CAT application and its implementation process.*

Keywords: Assessment or testing; computer adaptive testing; Item Response Theory; Predictive algorithm; Rubrics;

I. Introduction

For a long period, educational testing has focused mainly on paper and-pencil tests and performance assessments and has been used in different contexts, such as class presentations, essays, projects, practicum, etc.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues.

During the last two decades, in education, especially in the field of testing and training, usage of computers as a tool has spread rapidly in all levels. As a result, the computer-based testing (CBT) has increased significantly over the last few years. The most common type of CBTs is the linear and fixed-length assessments. The same number of questions is presented to each test-taker in a specific order and the score usually depends on the number of questions answered correctly. However, its use is limited, as there is a tendency to associate CBT with automated multiple-choice questions using multimedia and not as a tool that can enrich students' learning experiences [2]. Instead of giving each test-taker the same fixed test, CAT can adapts to the ability level of test-taker. There is a different way of selecting and presenting the items for each response of the test-taker. After each response it estimates the test-taker's ability and updates the item selection. For example, if the test-taker correctly answers the item presented, then a more difficult one is presented next. Otherwise, an easier item is presented next. Thus, low-ability test-takers will be presented with relatively easy items; high-ability ones will be presented difficult items. The calculation of the score for CAT will be based on the Item Response Theory (IRT) [2], [4], [7], [8]. In IRT, the test-taker can be described by a set of ability scores that are predictive, through mathematical modelling, linking actual performance on test items, item statistics, and test-taker's abilities. CAT is a useful tool in testing performance and shows promise in becoming one of the basic testing procedures especially in large-scale examination [2], [6].

CAT is still an open ended research problem, in terms of construction of a CAT as a number of parameters which are influenced a person's performance are involved such as utility, validity, satisfaction, usability, reporting, administration, security, and those associated with adaptivity and item pool. This paper is attempts to discuss the stages involved in developmental process with a generic framework of CAT application. In the

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues. JAB

following sections we have discussed about the framework and developmental stages of a CAT application.

II. Motivation

The Computer Adaptive Testing refers to those tests that are conducted online considering various parameters of adaptability. Many students across the world take computer Adaptive tests on various subjects of engineering, science, arts etc. In India also computer adaptive test is becoming crucial and is used for many examinations. The CAT is a software application which operates based on predictive algorithm that selects questions to the test taker from the item pool according to test taker's performance ability. Test taker has to attempt those questions and complete them in the stipulated time frame. Such model would be useful for all the learners in higher education spread across the country. At present nearly in all the institutions the tests are mostly paper based. Also, there are many demerits in the paper based test that CAT tends to take as an advantage. This application will have several benefits like:

- Human efforts can be reduced
- It can be carried out on a large scale.
- Time Reduction.
- The paper based test cannot be accessed remotely and also it is time taking.
- There is lot of manual intervention in paper based tests, need huge human resources.
- Selection of questions will dynamically according to level of the learner.

A. CAT is an important testing instrument in:

- Identifying whether the test-taker has met the specific objectives of a course
- Indicating the test taker's level of achievement in a skill domain

© 2012 Journal Anu Books Authors are responsible for any plagiarism issues.

- Identifying specific areas in which a student needs additional educational experiences
- Diagnosing the test-taker's skill area strengths and weaknesses
- Detecting whether candidates have met minimum course requirements as demonstrated in a mastery test [1].

III. Formalism and Architecture

While construction of a CAT, the developers must describe clearly the purpose of the assessment and they need to ensure that the test is able to measure the test-taker's true proficiency level. To reach this, a CAT should provide for a broad range of content areas and skill tasks - depending on the subject - to ensure that proficiency level is measured properly. Because different test-takers have different proficiency levels, the CAT has to be designed in such a way that it provides adequate assessment for the entire range of ability represented in the test-taker population [3] [5]. We generally find that while conducting a computerized tests, different set of approaches are acquired to deal with different contents. The test formula adopted for one particular content/course differs from the test formula adopted for any other course/content. There is an inevitable need for building a generalized formula that suits all the courses/contents. The test formula should be designed and formulated keeping in perspective of different levels of students in terms of their logical, psychological, comprehension abilities. Also the generalized test formula would be supportive to all kinds of contents.

The CAT would involve 4 set of people, they are Test maker and Test taker, Evaluator, Administrator. Their roles and responsibilities have to be defines across the three major stages which gradually defines a generic formula. Those 3 stages include:

- Before the test.
- During the test.
- After the test.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues.

A. *Stage-I: Before the test*

The computer adaptive test may face severe issues if crucial steps and precautions are not taken before its start. The Test Maker plays a major role in analysing the needs or requirements of the test (including Test Taker), gathering the data related to test and constructing an effective test. The tests are generally domain specific technical like C, Java, python) and non-technical like English, Telugu, social sciences etc. and are constructed in order to recognize or recall the answers.

The tests can be periodic and sequential. The periodic tests happen after regular interval of time where as the sequential tests are those which are interrelated or which follows a particular order. The tests can be conducted daily, weekly, monthly, after every module or any other basis.

1) Test Maker Responsibilities:

- Preparation and collection of item bank.
- Setting the type of questions and time factor.
- Preparation of rubrics i.e. evaluation parameters for correction of submissions.

2) Test Taker Responsibilities:

- The test taker has to go through mock test to reduce the fatigue and also to get habituation with the test pattern.

3) Administrator Responsibilities:

- The administrator has a crucial job of making ready all the logistics related to tests.
- The administrator also has to work on important factors like time, schedule of test and slot for the test.
- The administrator has to engage with test maker in order to do verifications before the test.

© 2012 Journal Anu Books Authors are responsible for any plagiarism issues.

B. Stage-II: During Test

1) Test Maker Responsibilities:

- The test maker has to make sure that all the test takers are able to get their user accounts and access the link/URL given to them.
- The test maker with help of administration should check and avoid any malpractice.

2) Test Taker Responsibilities:

- The test taker has to know his/her username and password along with the test URL.
- The test taker should be made aware of the pattern, subject and time.
- The test taker also has to know about the limitations of the software application which he/she uses.
- The test taker should avoid carrying any USB/electronic device to the test.

3) Administrator Responsibilities:

- The test maker should know the amount of hardware, internet bandwidth, backup, server and other technical issues needed for conducting test.
- There test maker should be able to solve the technical issues of test taker during the test.
- In case of any unwanted issue occurring like power – cuts etc, the test maker should be aware of its role.

C. Stage-III: After Test

1) Test Maker Responsibilities:

- Test maker has to make sure that the submissions in the server are equal to the number of test takers.
- Also there are no tampering done to the submissions and they reflect the exact marks.
- The test maker after the test should make an attempt to take a feedback from the test takers.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues.

- The test maker should take care of generation of results report with in specific period.

2) Test Taker Responsibilities:

- After the test and final submission have to check the result or report.
- Also need to provide a feedback to the test maker.

3) Evaluator Responsibilities [optional]:

- The role of evaluator arises if the questions given in the exam are descriptive type or the type of human intervention required.
- The evaluator should use rubrics to carry out evaluations.
- The evaluator has no role to play in case the questions are not descriptive. The evaluations in this case are carried out by the automatic evaluator.

4) Administrator Responsibilities:

- The administration should maintain privacy of the test results.
- The test takers feedback should be taken and analysed.

D. Criteria of Item Display

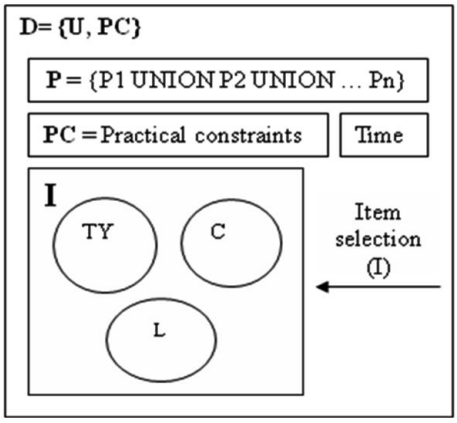


Fig. 1: Criteria of item display, where display of item (D) is the union (U) and practical constraints (PC). The Union (U) is presentation parameters (P), Time (t) and Item (I). Item is union of Type of Test (TY), Component (C) and Level (L) where rubrics (R1 ->L).

© 2012 Journal Anu Books Authors are responsible for any plagiarism issues.

The test maker has to fill the item pool with different types and items and levels. The items will be select by the predictive algorithm based on the item selection set. Selected items will be display to test taker with respect to above equation.

1) *Type of Test (TY):* It includes fill in the blanks, multiple choice, match the following, short answers, long answers etc.
2) *Components (C)*: The components are the domain-specific modules on which the test will be conducted. For example for Java content the modules can be Arrays, Strings, Threads, and Swings etc.
3) *Rubrics-I* (R1): This is an evaluation parameter with respect to time (t) taken by the TT during answer to the item.
4) *Level* (L): Level is the immediate response of the test taker to each question; it will be decided based on the R1->L.
5) *Time* (t): The time factor signifies the time needed for the test taker to answer to each question.
6) *Practical Constraints:* There are numerous practical constraints that may occur, CAT constructor has ready to be solved or mitigated when needed. Some of the major practical constraints have listed below; these constraints can be classified into technical, administrational, item-wise, test maker, test taker, application oriented, etc

- The insufficient internet bandwidth or slow internet speed if it is online.
- The number of computers needed for CAT with suitable configuration.
- The lack of power supply or frequent power cuts.
- Trained staff required in technical and administration.
- Operational and equipment costs are huge at initial.
- Server maintenance
  - The type of server and amount of load on the server.
  - Database maintenance also security threats to server.
- Security and privacy issues related to test taker.
  - The user accounts of the test takers should be safe.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues. JAB

- The site or the URL of the test should not get hacked.
- The test taker's performance is to be private, i.e. the results should not make notify to everyone.

- Habitational problem of test taker. The test takers must have gone through a mock test before the actual CAT. If possible a video clip can be shown to the test takers which tell how to attend CAT.
- A user friendly software application should be build and its must be modifiable and flexible.
- Application developers need to consider usability and navigational matters.
- etc

E. Block Diagram of CAT System

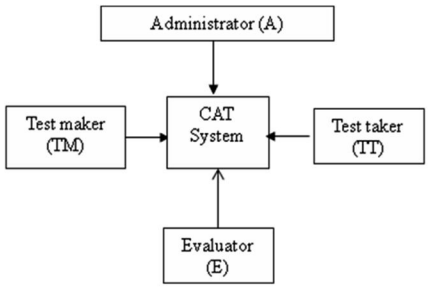


Fig. 2 This diagram showing the actors interacting with the system and the flow of data between them. Where the system is directly connected with actors are Administrator (A), Test Taker (TT), Test Maker (TM) and Evaluator (E).

The interface to interact with this system is dashboards of TT, TM and E respectively.

© 2012 Journal Anu Books Authors are responsible for any plagiarism issues.

1) *TM-Dashboard:* The test maker interacts with the TM dashboard and performs its activities. The TM-Dashboard interacts with the IRT where predictive algorithm, type of question and time factor are inputs and finally an item is selected. In this process TM-Dashboard performs the following:
   a. Passing parameters (TY, T, I) as inputs to the system for test.
   b. Setting parameters for evaluation.
   c. Final submission allocation to evaluators if it is descriptive tests.
   d. Checking feedbacks.
2) *Predictive Algorithm:* The role of predictive algorithm is very crucial in the CAT application. IRT will be part of in the functionality of this algorithm.

- The predictive algorithm interacts with the item pool database and selects an item based on the parameters given as input.
- Algorithm interacts with the previous levels of TT database and performs the storing and retrieving operations.
- This algorithm makes sure that the new question is displayed online which will be accessed by the test taker through the TT-Dashboard. It is a cyclic process as the algorithm continues taking inputs based on rubric and previous level of TT database and it continuously generates new questions by interacting with item pool database. The test taker submits the paper when the test is completed and the data goes into final submission database.
- The predictive algorithm needs a data structure as this algorithm interacts with the database to retrieve the information (previous level and item pool) and based on that it generates another level and using that level as a parameter it selects an item from the item pool. The Hash table data structure will be helpful in carrying out the operation.
- The steps involved in predictive algorithm have been given below.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues. JAB

- Steps involved algorithm:

Step-1: Input from TT (Test Taker), Input I/P= $Item_k$ (k ranges from 1 to n number of items given to TT)

Step-2: The input I/P are evaluated according to the rubrics.

The rubrics score is stored in a variable with name "rubric_score".

Step-3: The hash table (Ht) (it is derived based on a key-value pair concept) Ht_1 takes the evaluated score as input and generates related level integer using the Level Table in Ht_1

if key = rubric_score,

then

value = generated_Level.

The output level is stored in a variable with name generated_Level.

Step-4: If the item received from the TT is the first $Item_k$ such that k=1,

then

go to Step-5,

else

go to Step-7.

Step-5: Retrieve an item from the "Item Pool" database based on the "generated_score" value.

The "Item Pool" database has many Items of that "generated_level", so a random number program is used to select an item from the same level items.

The item retrieved is stored in the variable with name "new_item".

go to Step-6

© 2012 Journal Anu Books Authors are responsible for any plagiarism issues.

Step-6: The "new_item" is sent to TT (Test Taker/ Learner) who can visualize the question on dashboard screen.

If (n=total number of questions given to the TT/learner) then

go to Step-9

else

go to Step-1.

.i.e. All the questions are attempted by TT such that k = n.

Step-7: Retrieve the previous level from the "TT Previous-Level" data base and store the value in the variable "previous_Level".

$$\left(\frac{\text{Median (rubrics\_score, previous\_Level)}}{\text{Greatest(rubrics\_score, previous\_Level)}} \right)$$

P(new_Level)=

Where P is probability of new level

go to Step-8.

Step-8: Retrieve an item from the "Item Pool" database based on the "new_Level" value.

The "Item Pool" database has many items of that "new_level", so a random number program is used to select an item from the same level items.

The new_Level is stored in the "TT Previous-Level" data base.

The item retrieved is stored in the variable with name "new_item".

© 2012 Journal Anu Books Authors are responsible for any plagiarism issues. JAB

go to Step-6.

***Step-9*: END**

The below figure describes how the final result R is affected by the factors like type of question (TY), level (L), and time factor (t). Here q refers to the each selected question and *pl* refers to previous level.

$$Result(R) = \iint_{S} (L+t+TY)\partial q \partial pl$$

The factors TY, L and T are integrated with respect to *q* and *∂pl*. The entire CAT system can be denoted by S, such that:

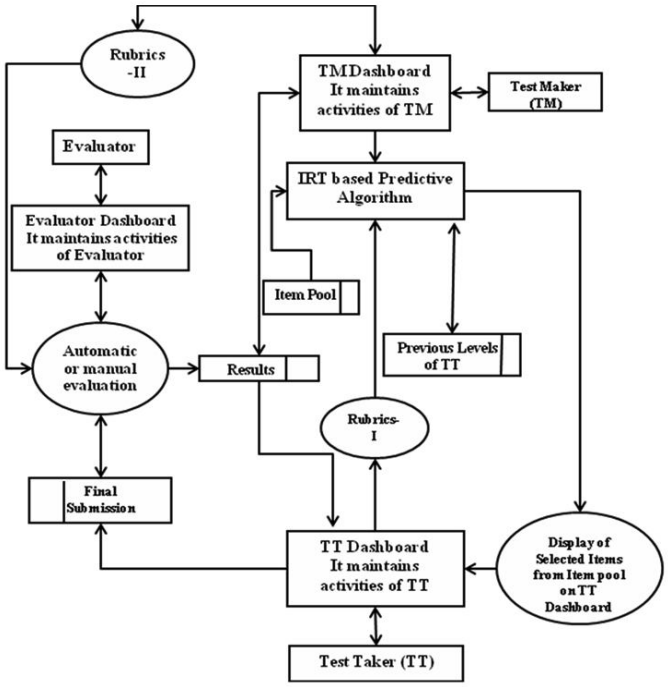


Fig. 3 Describes about the flow of data of Test Maker (TM), Test Taker (TT) and Evaluator (E) in a CAT model.

© 2012 Journal Anu Books Authors are responsible for any plagiarism issues.

3) *TT-Dashboard:* The selected item is displayed on interface, once the test taker logins in application with their login details. This item can be accessed by those test takers who have proper user credentials and permissions to take the test. The test taker (TT) gets its user credentials (username and password) by registration. Using that user account the test taker (TT) logs in and performs all its activities through TT-Dashboard which is meant specifically for TT. The TT-Dashboard has following role to play:
   a. Viewing the result/report generated in the end.
   b. Interface to write the online test.
   c. Feedback and comments.
4) *Rubrics-I:* The test taker attempts each question available in the online paper. The test taker can attempt the question (or skips it) and the TT-Dashboard takes each attempted question of the test taker and with the help of rubrics R1 (prepared by TM) the evaluation of the each question is performed and it finally generates the level.
5) *Rubrics-II:* The evaluator evaluates the final submissions based on rubrics-II and sends the scores to the result database. The result database can be accessed by test maker and the results from the result database can be viewed by the test taker through TT-Dashboard.
6) *E-Dashboard:* The next is the role of evaluator (E) who also interacts with E-Dashboard for performing all its online activities. The role of the E-Dashboard is following below:
   a. Access the answered paper or item sheet.
   b. Submit the results according to rubrics-II.
   c. Providing feedback.

   The evaluator has to evaluate the final submissions of the test taker. The evaluator also must register to get user credentials by administrator and using this user account evaluator can login into access the final submission through E-Dashboard. During this process the test maker

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues. JAB

distributes and allocates the existing submission of TT between the evaluators.

IV. Architecture and Process

The software application has followed a 3 layered architecture design. Here we have discussed the software engineering process and implementation details of CAT application architecture [9], [10].

A. Layered Architecture:

We have followed a 3 layered architecture with consideration of the following factors

- The client side where the learner fills the test sheet.
- Evaluator fills the results into the marks sheet.
- The data side where the data is processed and application server exists.
- The database where the learner entered data is stored.

Also the scores awarded by the mentors while doing the evaluation is also stored in the database.

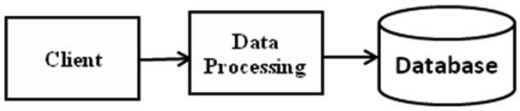


Fig 4 Three layered architecture design of the CAT system

This application is accessed by 3 types of users:

a. Learner or User

b. Administrator

b. Test maker

c. Evaluator

The application contains following pages:

- Home Page and Login
- Registration
- Dashboard (or My Account)

- Test Page
- Admin Page (Control Panel)
- Evaluator Page

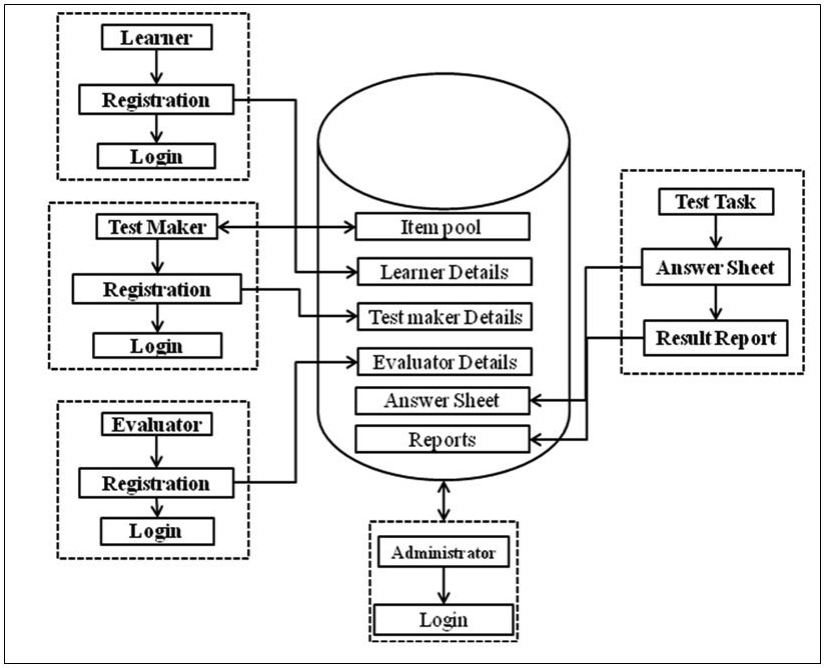


Fig. 5 The implemented architecture of online CAT application.

B. Modularity

While implementation the architecture processes have to divide into various modules or components. Each module or component plays a different role in the system as discussed below.

***Home Page*:** This page contains information about the application and general instructions, login window, forgot password link and other such options.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues. JAB

***Registration Page*:** All the users have to register before accessing the dashboard in this application. In this page the users have to give few details to maintain their credentials and the score.

***Dashboard or My Account Page*:** All the registered users can access their personal account page. All the details related tests will be displayed here with very brief information about their activities.

***Test Page*:** Once the user chosen or selected for writing the test, then s/he will be moved to the test page. Here the questions will be displayed with the options that you need to choose.

***Admin Page*:** This is a control panel to upload items into the database and some more facilities to change the application according to the user needs.

***Evaluator Page*:** The evaluator uses a set of parameters called rubrics to evaluate the answer sheets and allocate the marks. Those sheets with allocated marks are submitted by the evaluator when the evaluation is over and the submitted data is stored in the database.

***Predictive Algorithm*:** This algorithm predicts the level of the learner according to their answers given to the questions. Based on this level when the learner logins nextime to give test, it selects the item from the item pool with respective to the level. In the implimentation process we have taken the item respinse theory also to select the items as per level of the learner, the details have been described in above section. It is a propossed algortihm it requires to implimentation and also need to test the strength.

C. Information Hiding

Security and privacy are major functional aspects in this system. In this process the learners cannot access any database also they cannot access other learner's score reports to persistent the privacy.

Any evaluator also cannot access fully to control and to do any changes in the item pool also their role has defined to certain boundaries. Similarly, test makers also cannot access or change the learners' reports and

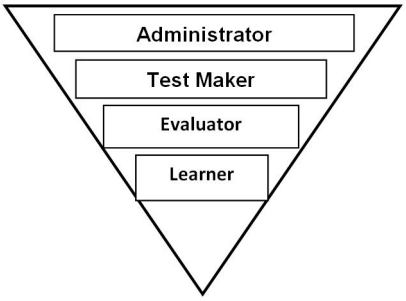
© 2012 Journal Anu Books Authors are responsible for any plagiarism issues. JAB

evaluators page unless permitted by the administrator. Test makers have been permitted strictly to manage the item pools.

D. Control Hierarchy

The control hierarchy is very crucial in any system; in view of CAT we have followed as the four main categories of users of this, those are Admin, Test Maker, Evaluator and Learner build the hierarchy of the system.

The hierarchy in the CAT system is follows:



E. Quality Attributes

Below are some of the quality attributes which are taken into consideration while designing and implimentation of the system.

***Scalability*:** It is as the number of learners and evaluators increase depending on the course and test requirements, the system database should be strong enough to handle such conditions.

***Extensibility*:** The CAT system has the capacity to deal with new changes at the architecture level. In fact it have a facility to collect feedback, based on the feedback in the future the system may see further modifications.

***Re*usability:** The system have always felxibility to add new features in it.

***Robustness*:** The technologies used to built the CAT system make it robust and it can quite well function in difficult conditions, such as low bandwidths.

***Security*:** The security of the system is given high priority, the databases cannot be accessed without permissions by the admin.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues.

***Fault tolerance*:** The available system is stable and can recover from any major failures which can occur at any level of the system. For example in databases and hirerarchies.

V. Implications

CAT is a form of computerized test that adapts to the learner level. For this reason, it has also been called tailored testing. There is a different way of selecting and presenting the items for each response of the test-taker. After each response it estimates the test-takers ability and updates the item selection.

Standard fixed tests almost always provide the best precision for test-takers of medium ability and increasingly poorer precision for test-takers with more extreme test scores. In contrast, computer adaptive tests can provide uniformly precise scores for most test-takers. So that, the test maker or a teacher can easily estimates the learners' level of understanding on the given item for testing.

This information will help to the teacher in further development of materials and methodology. This mechanism supports the results as privacy reporting, error analysis on errors done by the learner. This kind of reporting will reduce cognitive load also learner's inferiority complex.

VI. Future Work

This application required to implement the proposed predictive algorithm and also requires a greater number of perceptual tests with larger numbers of participating learners. The perceptual test results to be able to improve and validate it through a feedback process in the future.

It is also necessary to verify the property of adaptiveness of the algorithm comparing to formalism. This is therefore not the expression of a limited methodology, but rather one which is capable of evolving into new directions of future study.

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues. JAB

References

1. P. A. Dunkel, Considerations in developing and using computer-adaptive tests to assess second language proficiency. *Language Learning & Technology*, 1999, vol.2(2) pp. 77–93.
2. G.Elisabeth.T.Evangelos. & A.E.Anastasios Evaluation parameters for computer-adaptive testing, *British Journal of Educational Technology*, 2006, vol. 37(2), pp.261–278
3. B.Green, R. D.Bock, L.Humphreys, R.Linn, & M.Reckase, Technical guidelines for assessing computerized adaptive tests. Journal of Educational Measurement, 1984 vol.21, pp.347–360.
4. R. K Hambleton,, H.Swamination, & H.JRogers, H. J. *Fundamentals of item response theory*. California: Sage Publications Inc, 1991.
5. J.Harvey, & N.Mogey, Pragmatic issues when integrating technology into the assessment of students. In S. Brown, P. Race & J. Bull (Eds), *Computer-assisted assessment in higher education* (pp. 7–20). London: Kogan-Page. 1999
6. M.Lilley, & T.Barker, *The development and evaluation of a computer-adaptive testing application for English language*. In Proceedings of the 6th Computer-assisted Assessment Conference. Loughborough University, UK. 2002.
7. F.M.Lord, F. M. *Applications of item response theory to practical testing problems*. NJ: Lawrence Erlbaum Associates, Publishers, 1980.
8. H.Wainer, Computerized adaptive testing (a primer), NJ: Lawrence Erlbaum Associates, 1990.
9. I. Sommerville, Software Engineering, 7th ed, Addison-Wesley, 2004.
10. R.S.Pressman, Software Engineering-A Practitioner's Approach-Required, McGraw Hill, 1992.

* * * * *

© *2012 Journal Anu Books* Authors are responsible for any plagiarism issues.