ONLINE ADAPTIVE ASSESSENT PLATFORM

TECHNOLOGY BUCKET: Software-Web App development **CATEGORY:** Software

COMPANY NAME: SSEPL Skills Pvt Ltd. **PROBLEM CODE:** RA2

TEAM LEADER NAME: Abhishek Paharva **COLLEGE CODE:** U-0564

Table Of Content:

1. ABSTRACT

Computerized Adaptive Testing (CAT) is a form of computer-based testing. It adapts to the ability level of the examinee and hence also referred as Tailored testing. CAT can reduce testing time 50-90% with no loss of precision measurement. They are a sophisticated method of test delivery based on the Item Response Theory(IRT). They operate by adapting both the difficulty and quantity of item seen by each examinee. In the last few years, CAT has been adopted for formative and summative assessment activities in educational field and in online learning in particular. While a variety of platforms are available for designing and deploying CAT but the real challenge of providing the flexibility in test and item design that is required for domain specific assessment formats in education has hardly been addressed. Developing a sound, defensible CAT is not an easy task, but our goal is to make it as easy as possible- that is , everything you need is available in clean and flexible software UI and also introduce a well-defined software architecture to enable the development of fully customizable CAT tools with respect to domain specific item design and visualization. The prototypical architecture outlines the software, data analysis, and project management steps needed to develop a computerized adaptive test that aligns with best practices and standards.

KEYWORDS

Computerized Adaptive Testing , Tailored testing, Item Response Theory

2. INTRODUCTION

Computerized Adaptive Testing (CAT) has been applied in educational psychology. It is a form of computer administered test in which the next item or set of items selected to be administered depends on the correctness of the test taker's responses to the most recent item administered.

There is a long history of using adaptive testing for improving the test results. The test administrator chose questions depending on the response of the candidate. Because having an administrator available for each candidate was costly, these adaptive tests were never used on a large scale. However, such a testing program is still useful to consider as an ideal test, where each candidate receives an exam administered specifically for that candidate. Computer adaptive testing aims to move closer to that ideal.

Currently available CAT systems focus on items that can be answered dichotomously or on a multi-part scale.CAT is primarily used in combination with IRT models Items prepared for use in IRT can take different forms. If an item can be evaluated unambiguously to be either true or false, it is called "dichotomous". If there are more than two response options, the items are referred to as "polytomous". While this is appropriate for latent-trait-testing, the primary use-case of CAT in psychometrics, the evaluation of learning outcomes might requires complex and open answering options.The items requiring the presentation and

evaluation of answers comprising of multiple components (e.g. several input fields, offline test access,locating multiple errors in conceptual drawings such as electronic circuits or process models) are not supported. Also, the items containing dynamic elements that rely on user interaction (such as assessing the behaviour of the physical system) can only be administered when using third-party technological solutions.

The present work addresses the above limitations and aims at providing a flexible and interactive architecture by putting a particular focus on integration in online learning settings. The architecture is designed in a way that does not only allow to alter the types of items, but also the testing strategies, the algorithms for item selection, and the user interface. Furthermore, this architecture provides features including question bank, candidate management facility, test creation and reporting system. It also includes Offline assessment environment to eliminate barrier of internet connectivity. It also includes displaying of Individual Ranking of exam result, pass/fail status , topicwise performance of candidate and questionwise performance of the whole candidate slot.

Using the proposed architecture enables technology-proficient users to integrate CAT in their online learning platform and provides a light-weight, XML-based, item specification format to domain experts responsible for maintaining the item pool.

3. COMPUTRIZED ADAPTIVE TESTING FOR ONLINE LEARNING

CAT successively selects questions with the aimto maximize the precision of the exam based on what is known about the examinee from previous questions. For the examinee's, the difficulty of the exam is tailored itself to their level of ability. For example, if an examinee performs well on an item of current difficulty level, they will then be presented with a more difficult question, that is their level of difficulty will be upgraded. Or, if they performed poorly, they would be presented with a comparatively simpler question, that is their level of difficulty will be degraded.

Compared to static multiple choice tests which have a fixed set of items administered to all examinees, computer-adaptive tests require fewer test items to arrive at equally accurate scores.

In computer adaptive testing, the difficulty level of the test items is determined by the ratio of the number of past candidates who answered the item incorrectly to the total number of candidates who viewed the item. An item that many candidates get incorrect is determined to be difficult. An item that many candidates get correct is determined to be easy. A candidate who answers correctly items that many candidates answer incorrectly will get a higher score than a candidate who answers correctly those items that nearly all candidates answer correctly. Using computer adaptive testing models, there is no subjective measure of an items difficulty. Difficulty is strictly a statistical parameter.

4. ITEM RESPONSE THEORY(IRT)

Item Response Theory (IRT) represents a significant innovation in the field of psychometrics. The foundation of IRT is a mathematical model defined by item parameters.

It is a paradigm for the design, analysis, and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes, or other variables. It is a testing theorybased on the relationship between individuals performances on a test item and the test takers levels of performance on an overall measure of the ability that item was designed to measure. It can be used with a variety of item selection algorithms and scoring procedures. They try to estimate an examinee's skill level and therefore find a relation between an examinee's answers to particular items and their skill level.

There are different approaches for this purpose. The most widely used approach is the 1-parameter-logistics-model that only requires to measure the difficulty level of each item. The remaining parameters- discrimination, that is, the amount of information the item provides for skill estimation, and guessing, that is , the probability of the guessing the right answer- remains fixed. 2- and 3-parameter-logistics give a more exact and faster estimation of an examinee's skill level, but require to determine the additional parameters for each item.

5. CAT PROCESS

The administration procedure of a computerizedadaptive test (CAT) can be broadly described as follows:

1. The computer selects an item from a large collection of items (item bank).
2. The item is presented to the candidate on the computer screen.
3. The student responds to the question by typing or selecting an answer.
4. The computer evaluates the response as correct or incorrect.
5. If the answer was wrong, the next item presented will be easier; a more difficult item will be administered in the case of a correct.
6. The computer terminates testing when some stopping rule is satisfied.

Besides a computer, the process of adapting a test to an examinee requires an item bank and a testing algorithm.

A. Calibrated Item Bank

Adaptive testing requires a calibrated item bank. An item bank is a large collection of accessible test items. Accessible means that the items are classified or organized in such a way that they can be retrieved easily for test assembly. The items in an item bank are usually classified according to content, question type, performance type, cross-reference to other items or to common stimulus material, author, testing history, and psychometric characteristics, including the difficulty level. Item response theory solves the problem of classic test theory by providing estimates of difficulty parameters that are not dependent on the ability level of the group upon which they were developed. Although, in IRT, the

difficulty parameter of a particular item is group-independent, the chances of giving a correct answer are different for individuals: the high-ability person's chances of giving a correct answer will be higher than those of the low-ability person. Sample-independent estimates of item difficulties are possible because, in IRT, the interaction between the ability of a person and the difficulty of an item is explicitly modeled. There are several models which describe this interaction. When a particular model is chosen, the tenability of it is investigated in a so-called calibration study involving large samples of examinees. Items which do not 'behave' as predicted by the model must be deleted. The remaining items have difficulty values which are valid for every prospective examinee in the population from which the sample was taken. These items can be rank-ordered according to their difficulties on a scale. The scale concept implies that an examinee who gives a correct answer to a particular item will probably also give correct answers to items with lower scale values and wrong answers to items with higher scale values. Similar scales can also be built based on classical test theory, but these scales would be very unstable in that they would fluctuate with changes in the group whose responses were used for establishing item difficulties. An essential characteristic of IRT is that a person's ability can be expressed on the same scale as the item difficulties. The concept of the common scale is very helpful in understanding the principle of adaptive testing which says that the next item to be administered in an adaptive test has a difficulty value which is as close as possible to the estimated ability of the person.

B. Testing Algorithm

The testing algorithm is the heart of an adaptive test. Its specifications determine how the test must be started, continued, and stopped.

Starting:In an adaptive test, the difficulty of the next item to be administered is determined by the current estimate of the examinee's ability. Each time the examinee gives an answer to an item, his or her ability is re-estimated on the basis of all the answers given so far. However, when testing begins, information about the ability of an examinee is often lacking. Many CATs solve this problem by selecting an item with a difficulty level that is adapted to the average ability of the testees in the calibration study.

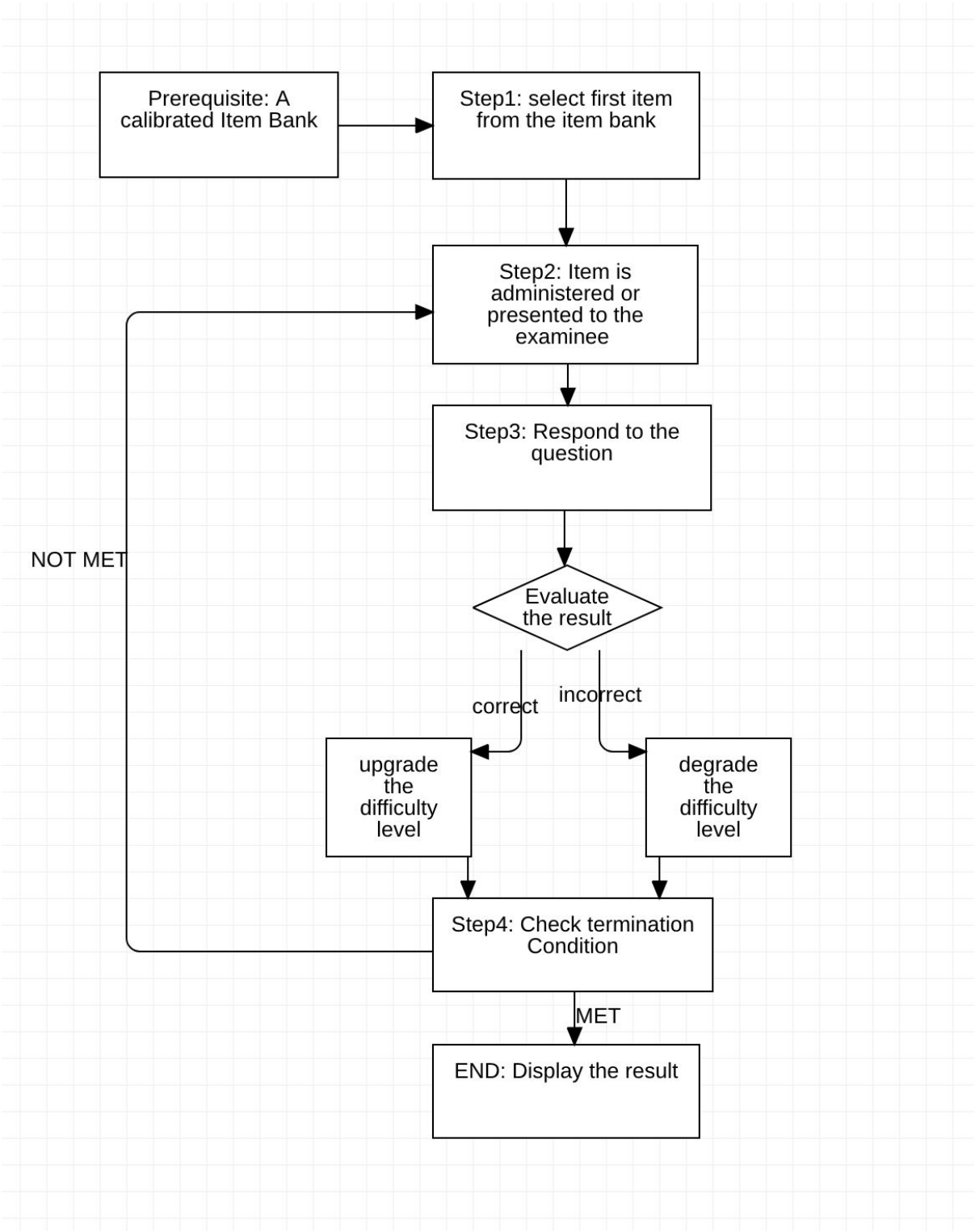
Continuing:The main reason for the existence of adaptive tests is their efficiency. Efficiency in testing means that fewer items are needed to estimate the abilities of examinees with a specified level of accuracy or for taking decisions about persons. Test A, for instance, is more efficient than test B if it has more or less the same measurement accuracy or leads to the same decisions about persons but needs fewer items to accomplish this. The relationship between the concepts of measurement accuracy and information explains why the emphasis in the item selection process of many CATs is on maximum information. The maximum information selection strategy implies that the next item to be selected is the one which provides the most information about the latest ability estimate of the examinee. In general, this will be an item which has a difficulty that matches the ability level of the examinee.

Stopping:An adaptive test is terminated when a particular stopping rule has been satisfied. The stopping procedure is always a combination of psychometric and more practical rules so that the test can be stopped in case the psychometric criterion cannot be met for some reason. The more well-known practical stopping rules are fixed test length and fixed testing time. Psychometric stopping rules are those which are defined in terms of measurement precision. Basically, there are two ways to evaluate measurement precision. In the first, testing is stopped when the standard error of the latest ability estimate is equal to or smaller than a certain criterion value. In the second, no criterion value for the standard error is defined. Instead, a cut-off score is defined on the ability scale which separates masters and non-masters. The test is stopped when the cut-off score falls outside a confidence band around the examinee's ability estimate.

6. CHEATING PREVENTION TECHNIQUES:

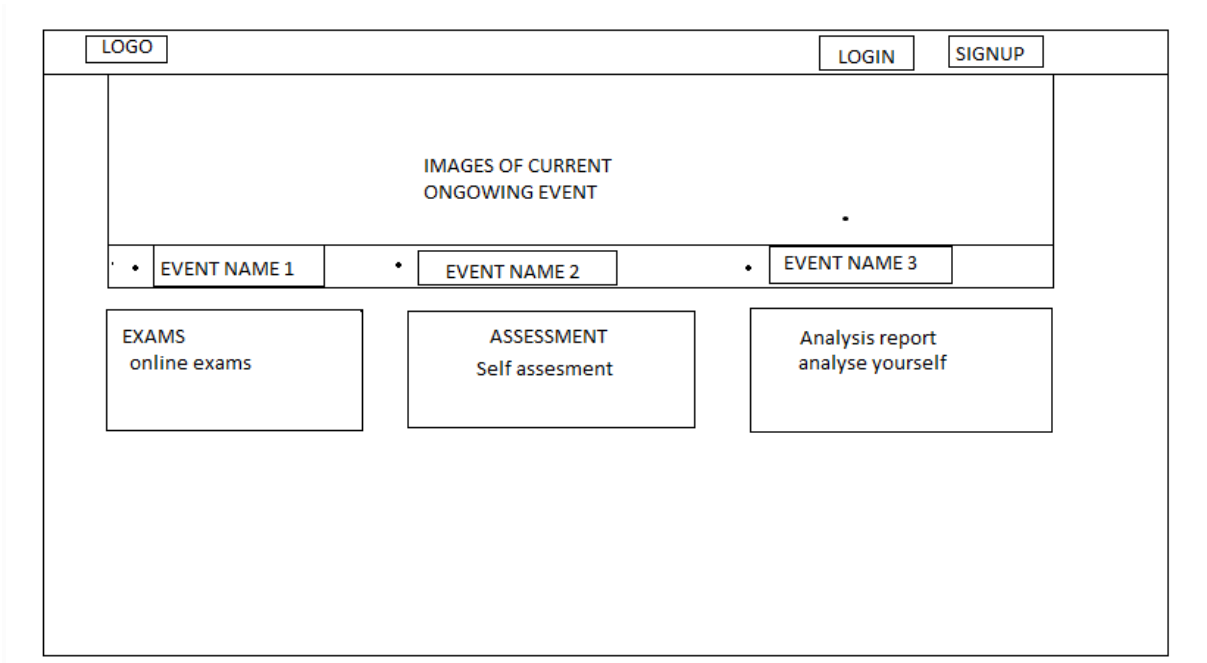
A. Window procting:
   The test window is proctored if the candidate navigates away from the test window (for instance to use google search), multiple warnings are given. If the warning ignored the test is ended immediately and the red flag is raised in the report.
B. Copy detection:
   When code is submitted by any two candidates are too similar a red flag is raised indicating possible code plagrism.
C. Random pooling:
   Each candidate giving test are provided with smaller set of question that are randomly selected from large pool of question.
D. Question shuffling:
   Each student giving test are provided random question from large pool of questions.
E. Option shuffling:
   Option provided to the student for each question are shuffled.
F. Multiple login is not allowed:
   One student cannot login same test from multiple screens.

7. FLOWCHART ASSESSMENT PROCESS

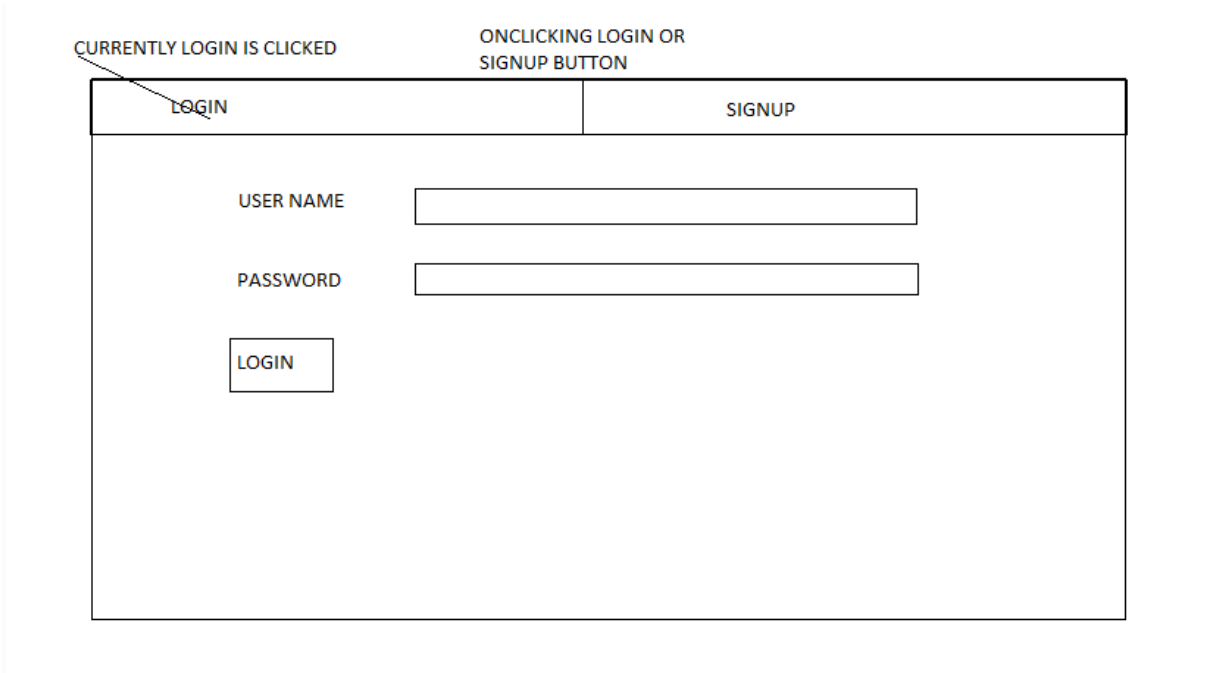


Flow-Chart. Assessment Process

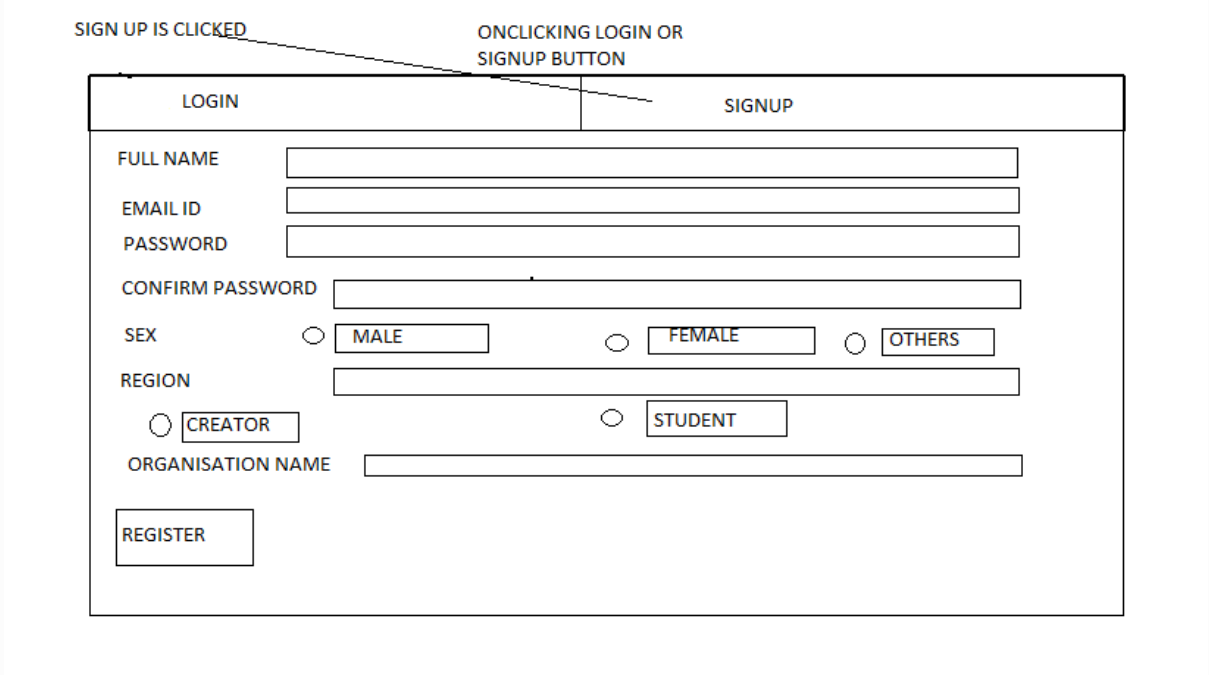
8. GUI PROTOTYPE :



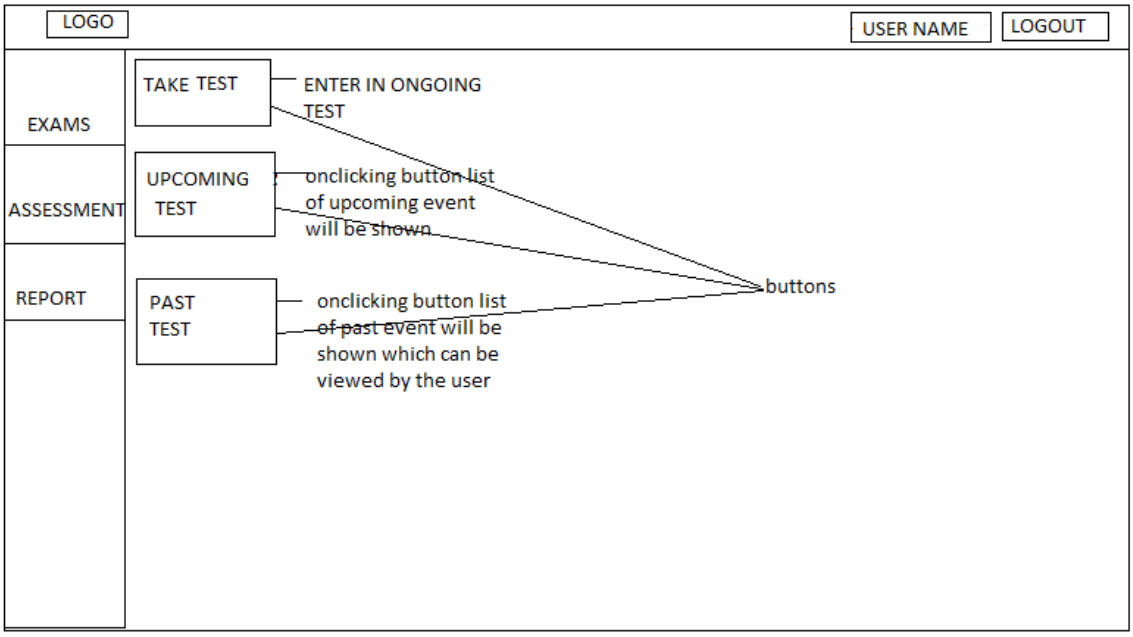
1. Index Page



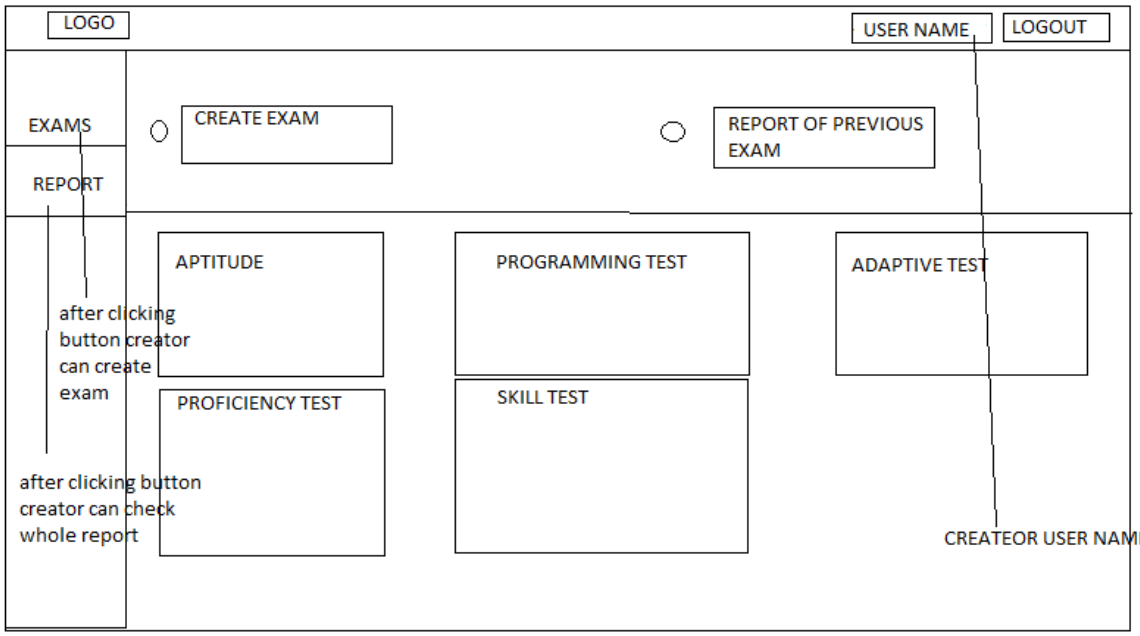
2. Login Page



3. Sign-Up Page

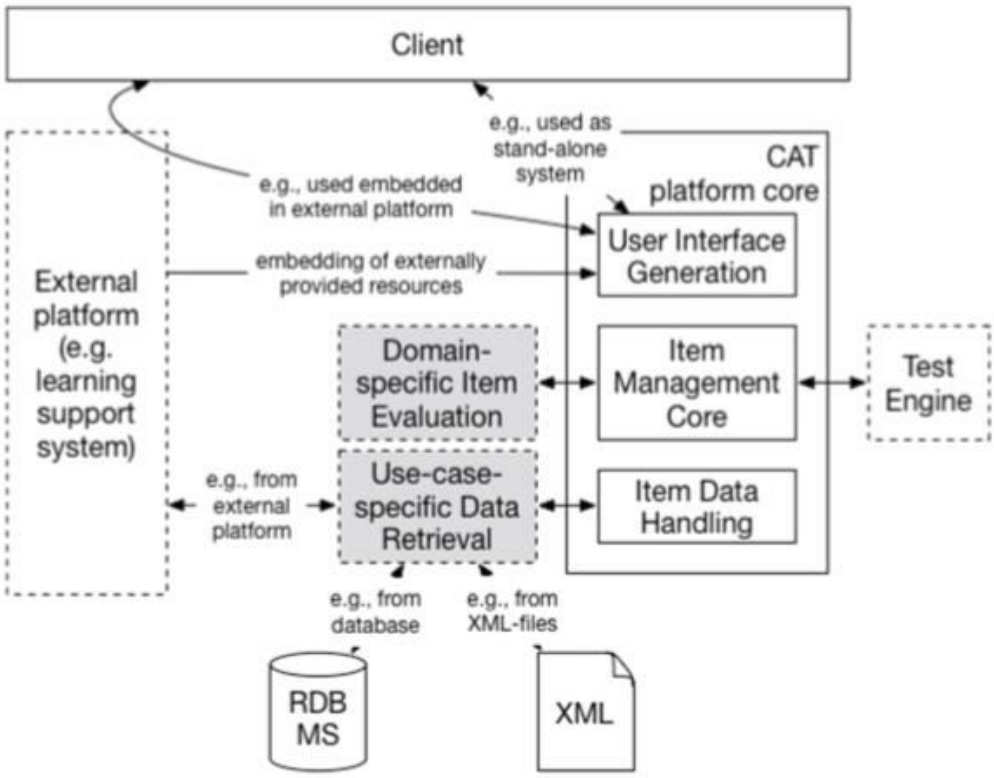


4. Student Dashboard



5. Creator Page

9. PLATFORM ARCHITECTURE :



10. TECHNOLOGY STACK:

Client Side:

- **HTML, CSS, Bootstrap, JavaScript,**

Server Side:

- **PHP**

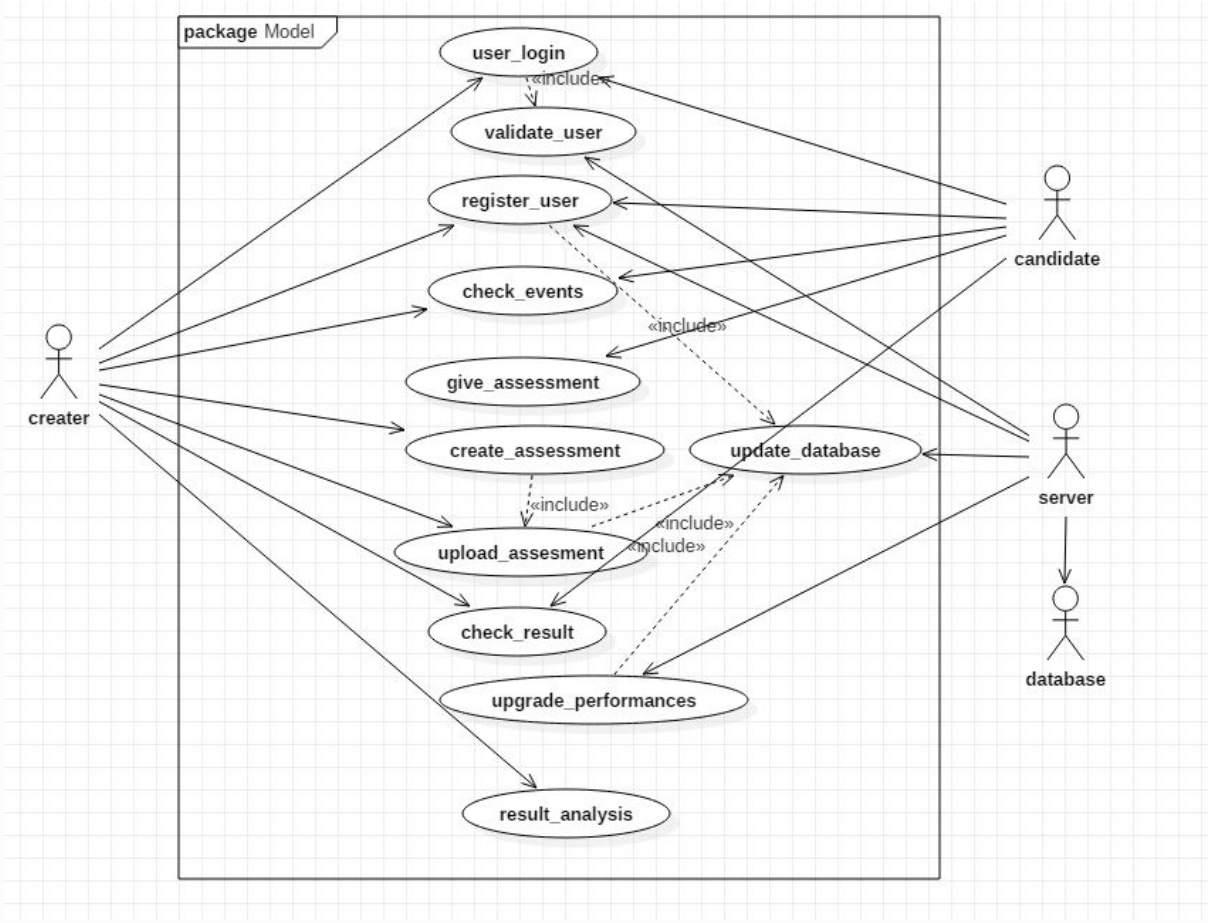
Database:

- **MySQL**

Server:

- **Apache**

11. USE CASE DIAGRAM



12. IMPORTANT FEATURES:

- **Platform will be dynamic and reliable as per the needs of the required user.**
- **Proper session management and injection attacks prevention will be provided as far as the security is concerned.**
- **Cheating prevention techniques are used so as to make the assesment more neutral i.e. biasing is exempted of any kind.**