International Journal of Recent Research and Review, Vol. IX, Issue 1, March 2016
ISSN 2277 – 8322

A Model for Automation of Subjective Assessment in Online Education

Deepnanda Ray, Piyali Chakraborty
School of Education Technology, Jadavpur University, Kolkata, India
School of Education Technology, Jadavpur University, Kolkata, India
Email: deepnanda.ray@gmail.com

***Abstract* -The most popular online assessment techniques in the objective domain are Multiple Choice Questions, True/False, Fill in the Blanks or Labelling a Diagram.A major reason for their popularity is ease of automation programming for these assessment techniques. Once these questions are uploaded in the system as part of a question bank, along with relevant correct options that should be selected as answers, the system can carry out the assessment on its own. A simple randomize algorithm can further enhance the unpredictability of this question bank. However, such techniques are more suited to a formative assessment than a summative assessment. Summative assessment is targeted towards knowledge and comprehensibility of the student at the end of the course and hence many teachers feel the necessity of including some subjective questions in addition to objective types to test the students. This is particularly true for subjects of some disciplines like humanities where writing a passage of text might be absolutely indispensable to understand the command of the student on the particular subject.This poses a serious challenge to the idea of automated education in an online mode. Such subjective assignments cannot be evaluated automatically. Substantial time and effort from the concerned teacher would be required for exhaustive analysis and marking of subjective answers. Indirectly this might pose other constraints on the online system, e.g. the number of students who can enrol for the course. Options like self-paced learning would not be applicable, because assessment cannot be automated entirely. In this paper a 5 step intuitive solution to this issue has been proposed.The first two steps of Grammatical Correction and Spell Check act as filters and take the input text towards a canonical format. At the same time the data on the number of aberrations that are identified during a particular step is maintained in a database and this data contributes to final step inConsolidation of the Marks.Once the input text is taken to a version free from grammatical errors and spelling mistakes (step 1 & 2) , a third step of keyword/key phrase search is conducted on this text and percentage of match is maintained in a database and this data contributes to the final step in consolidation of the marks. For descriptive answers that explain a certain process (e.g. photosynthesis) or some chronological episode (e.g. a historical event), we can use an optional step 4, where an abstract, i.e. a model answer in précis format is input. As an additional input, the expected cut off percentage of accuracy can be entered, e.g. x%. If x% of the abstract cannot be identified within the input text, the answer is declared to be theoretically insufficient. Any match above x% is maintained in a database and this data contributes to Step 5 in calculating the final marks on the basis of a well-defined algorithm that is customizable for each topic/subtopic/individual question.**

***Keywords*– humanities, subjective assessment, automation.**

I. INTRODUCTION

The research is focussed on a novel approach of assessment that can handle subjective responses. In order to achieve this, we have referenced several very well-known tools, such as tools for spell-check, grammar check, word matching, substring count and text comparison. Some of these tools were customized slightly to map to the research requirements. Such tools will be discussed in details in the section Customization of Existing Tools.

Also, a detailed literature survey was done to understand the underlying philosophy of assessment. Assessment is viewed as “the process of gathering and evaluating information on what students know,

understand, and can do in order to make an informed decision about next steps in the educational process" [1]. The goal of formative assessment is to monitor a learner during the learning process. This is called "test run" process. Through this process a teacher can recognize student's improvement. From the ongoing feedback the teacher can modify his/her teaching style accordingly. Black *et al.* [2] gave a detailed description of formative assessment for learning: *"Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of modifying student behaviour, or of certifying competence. An assessment activity can help learning if it provides information that teachers and their students can use as feedback in assessing themselves and one another and in modifying the teaching and learning activities in which they are engaged. Such assessment becomes "formative assessment" when the evidence is actually used to adapt the teaching work to meet learning needs" (p. 10).* From this definition it is clear that feedback is the core part of formative assessment. Formative assessment helps teachers determine next steps during the learning process as the instruction approaches the summative assessment of student learning. On contrast of summative assessment, formative assessment gives a qualitative feedback rather than overall result. Summative assessment focuses on the final outcome of an instructional course. Its leads to traditional assessment [3] i.e. it summarises the learner's achievement [4] at the end of a course or session or at the entry time of college. A teacher assesses a student at the end of a course against a standard benchmark. A comprehensive assessment program at the classroom level balances formative and summative student learning information [5]. The Partnership for Assessment of Readiness for College and Careers (PARCC) recognizes that teachers "need additional support to collect evidence of learning to inform instruction, hour by hour, day by day, and week by week" [6]. As Shepard [7] has noted, "formal theory about formative assessment was developed in other countries [8-11], in part to counter the negative effects of external accountability tests exported by the U.S." (p. 2). Harry Torrance and John Pryor commented that "formative assessment per se, as opposed to formative assessment distinguished from summative assessment, has received relatively little attention"[12] .Current research extends focuses on formative/summative assessment is the alternative method of evaluating learning. Beginning from a systems perspective conceptualization of feedback as "information about the gap between the actual level and the reference level of a system parameter that is used to alter the gap in some way" [13]. Sadler accepted that the formative assessment as a feedback loop to close the gap between the learner's current status and desired goals [14]. Some researchers (Bloxham& Boyd) [15] propose four purposes namely certification, quality assurance, student learning and lifelong learning capacity. Some reduce this to three including assessment of learning, assessment for learning and assessment as learning [16].

II. REQUIREMENT ANALYSIS AND DESIGN APPROACH

The online subjective assessment module, discussed in this section, is designed primarily for summative assessment of the student. This software can be made available in a plugin form that can be integrated to any online learning system. This system should ideally be designed such that there would not be any necessity for offline review/input, once the initial database is populated. The entire process should be implicitly automated to minimize the time requirement from the teacher/instructor. The students should be able to type in their assignment onscreen or upload a file to the system. The system should be modal in nature, i.e. no other program can be accessed when this system is running. This is required in order to avoid chances of cheating/referencing.

The keywords/key phrases and the abstract for each individual question need to be populated by the teacher/teaching assistant. There might be substantial effort required initially to prepare this. However in the long run, a question bank would be available for the system along with a corresponding abstract and keywords/key phrases for each question.

During the research phase a survey was done among undergraduate students on 3 basic questions from 3 different subjects. The questions were:

1. How can electromagnetic force are generated in a magnetic field. Explain.
2. Write a short note on Photosynthesis
3. Write a short note on the evolution of armed nationalist movement in Bengal in pre-independence era.

The answers were studied and sampled in order to arrive at the keywords, key phrases and the abstract corresponding to each question.

E.g. after a thorough study of the different responses, the following keywords, key phrases and abstract were identified for the question "Write a short note on Photosynthesis":

Keywords:

Chlorophyll, Carbondioxide, Oxygen, Green, Sunlight, Glucose

Keyphrases:

"Photo = Light, Synthesis = Combination"
"$6CO_2$ (Carbondioxide) + $6H_2O$ (Water) ----light----> $C_6H_{12}O_6$ (Glucose) + 60_2 (Oxygen)"
Abstract:
Photosynthesis is a process in which the green plants prepare their food. Green plant contains green pigment called chlorophyll which traps solar energy and convert it to chemical energy. In the chemical reaction of photosynthesis carbondioxide and water combine in presence of sunlight and form glucose and oxygen. The oxygen is released in the air and glucose stores in the form of starch in leaves and other part of the plants.

This method could be followed in real life, or the teacher/teaching assistant could create this data by themselves. However, a sampling of student data gives an idea of probable variance in the answers that can be expected, and this also helps in identifying the cut off percentage of accuracy.

The software design was done with three major components: A database schema called "Assessment Data Bank" from which the keywords/key phrases and the abstract corresponding to a particular question is retrieved, a process called "Evaluation" which has business logic that matches the entered answer with these components, apart from doing the regular spelling and grammar check, and another process called "Result Generation" that works in tandem with the "Evaluation" module to generate marks. The "Result Generation" process updates a "Student Marks" schema, where marks calculation is done individually for each student based on their performances.

The following flow chart explains this further:

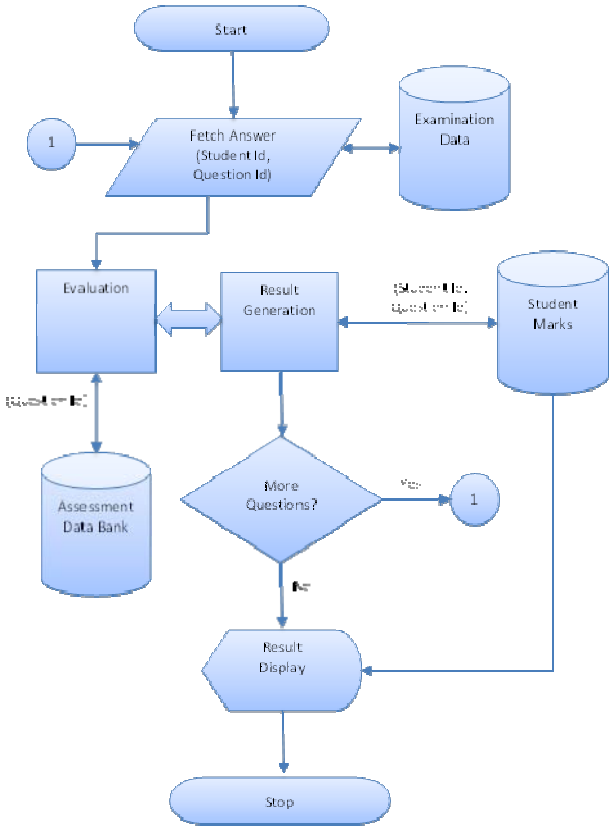

Fig. 1. System flow chart

The answer corresponding to each question ID in a particular test is fetched from the examination database, corresponding to the student ID. This answer is analysed by the Evaluation process, using the data (keywords, key phrases, abstract) from the Assessment Data Bank and the Result Generation module calculates the corresponding result. The process is iterated for each question in the test, and finally the consolidated

result is displayed. The Section “Business Logic “explains further about the logic involved in the Evaluation and Result Generation modules.

III. CUSTOMIZATION OF EXISTING TOOL

The working principles of the referred tools are explained along with screen shots. The textual input for this phase of research was generated from the survey that is discussed in the section Requirement Analysis and Design Approach.

The user interface for spell and grammar check tool that was referred [17]:

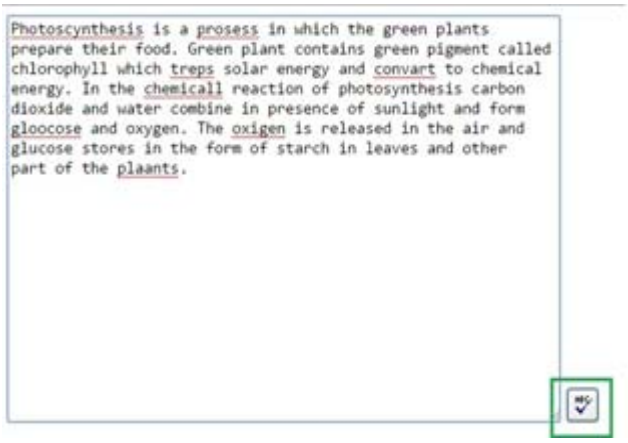

Fig. 2. Screen shot showing the initial input screen

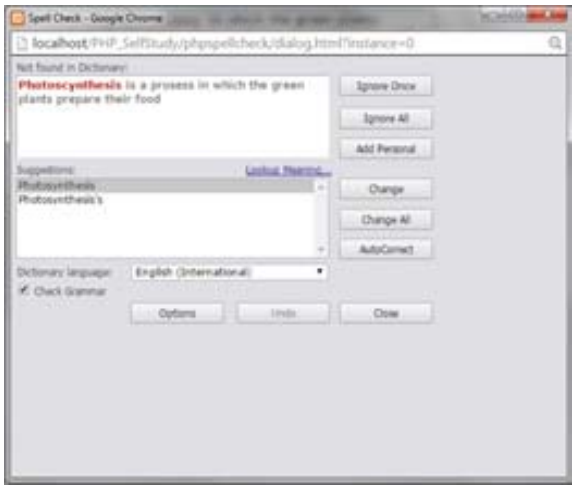

Fig. 3. Screen shot showing the spell and grammar check pop up

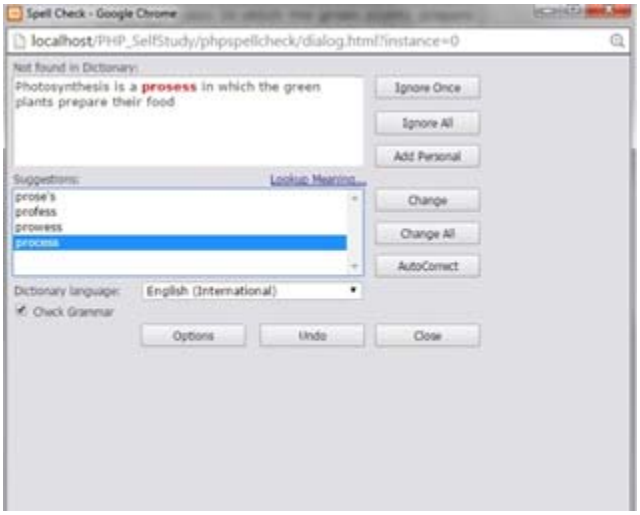

Fig. 3. Screen shot showing predictive word for reference

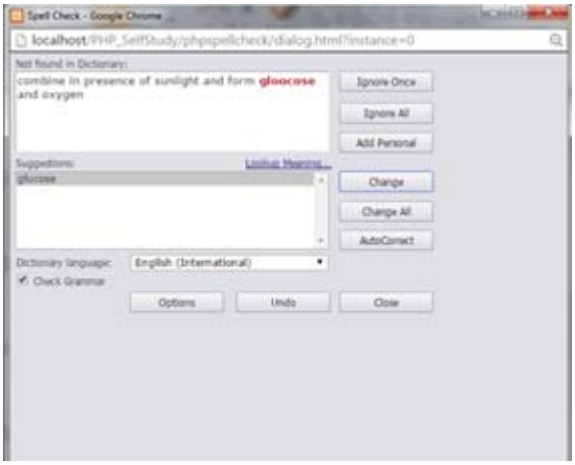

Fig. 4. Screen shot showing predictive word for reference

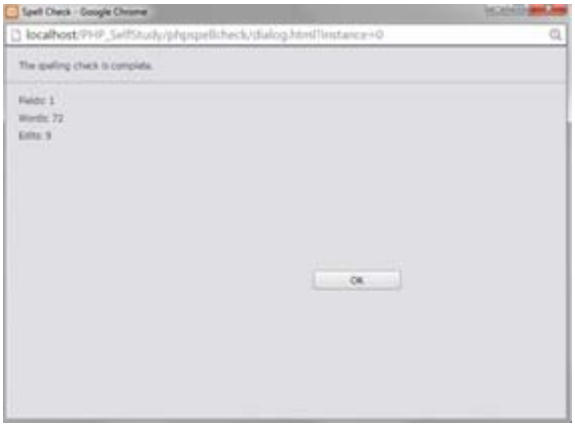

Fig. 5. Screen shot the results for spellcheck

The result screen shows the number of fields (paragraphs), total number of words content in the input paragraph and the number of errors. From this statistics the result is calculated. For example from the above statistics out of 72 words 9 errors are found. So the percentage marks is calculated of the basis of ratio of total word count and errors found (9/72 * Credit marks for spell and grammar check). This tool not only finds out the mistakes but also texts the input text to a canonical form to facilitate Keyword/Keyphrase/Abstract matching.

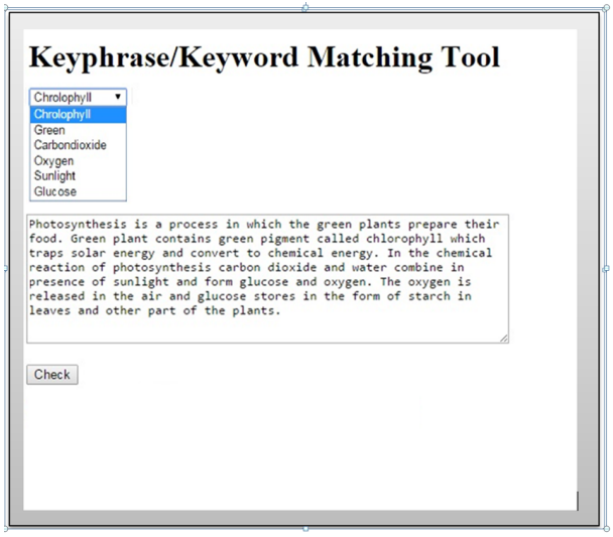

Fig. 6. Screen shot showing the key phrase matching interface

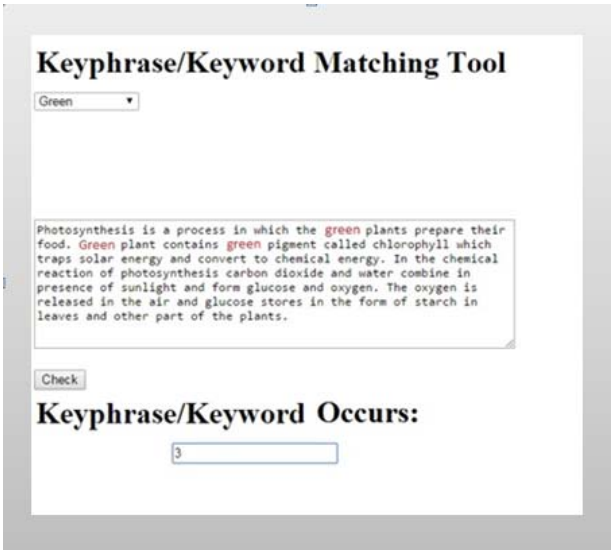

Fig. 7. Screen showing the key phrase matching result

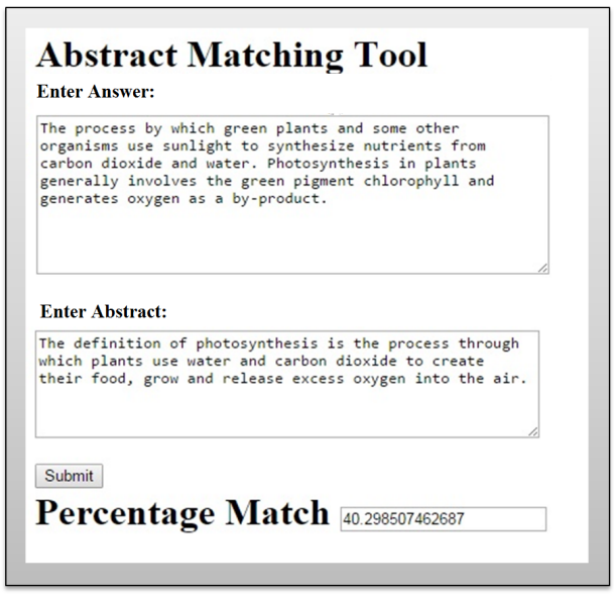

Fig. 8. Screen shot showing the abstract matching result

These tools have contributed to the Interface Design and form the basic components of the Evaluation Dashboard.

IV. INTERFACE DESIGN

Although the initial aim was to automate the process, the interface is planned in a way, such that the Evaluation process could be monitored by a teacher or a teaching assistant. This was done primarily because the research is still in its initial phase, and a close monitoring is required to identify probable errors and omissions, so as to enable further betterment of the system. The interface is explained briefly with the help of diagrams.

The student logs in to the Test System and enters the answer set, either by typing or by submitting an attachment. The answers are parsed, categorized and stored in an Examination Database, against Test ID, Student ID and Question ID. Let us assume that Test ID 1 has 4 questions, and Student ID 05 has appeared for the test and submitted the answers. Our system can be accessed for evaluation at this stage.

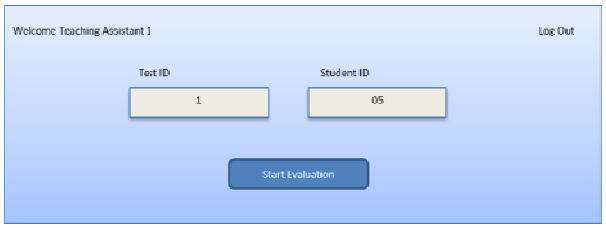

Fig. 9. Screen 1

The initial screen can be accessed by appropriate roles (teacher/teaching assistant) by clicking on the plugin link in the online learning platform. The Test ID and the Student ID need to be entered, so that the corresponding answers are loaded into the memory from the Examination Database.

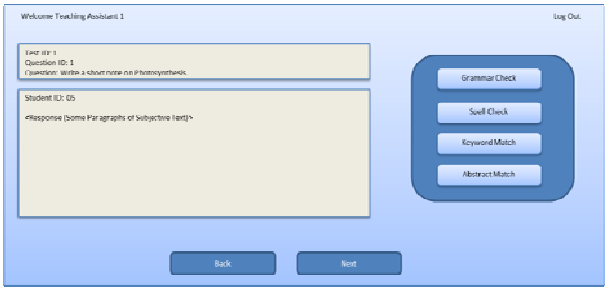

Fig. 10. Screen 2

The answer corresponding to the first question (Question ID 1) of Test ID 1, as submitted by Student ID 05, is displayed on screen. Towards the right of the screen is the Evaluation Dashboard, comprising of 4 toggle buttons which are sequentially activated for each stage of evaluation. On click of any of these buttons, a pop up screen appears, to show the detailed result of that particular stage, e.g. “Grammatical Mistakes - 7”, “Spelling Mistakes - 3”, “Abstract Match – 73.5%” etc. For Keyword/Key Phrase matching, each keyword as obtained from the Assessment Data Bank is matched against the answer string and a tabular synopsis is provided.

Table 1
List of keyword occurrence

Keyword	Number of Occurrence
Sample1	3
Sample2	1
Sample3	0

On click of the Next button, the system moves on to the next question-answer set, i.e. the answer provided by Student ID 05, for Question ID 2 in Test ID 1.

Fig. 11. Screen 3

This screen is accessible once all the questions in a particular test are evaluated. This screen displays the consolidated result for the particular Student ID. Marks for each individual question are shown, along with the total marks. Moreover, in the remark column, remarks are displayed against specific cases, e.g. “Did not attempt” in case no entry is found against a particular question, and “No Match Found” if there is not a single instance of matching keyword, key phrase or abstract. In both these cases the marks is considered as 0. A column is provided to display the details against each question, e.g. details on the evaluation logic, and marks deducted at each phase of checking (this is detailed in the Section “Business Logic”).

V. BUSSINESS LOGIC

The business logic comprises of the actual assessment logic and the weightage associated with each phase of evaluation, the logic involved in calculating the total marks, and a subjectivity parameter associated with individual traits of each student, which helps in measuring their level of improvement with respect to their historical performances in similar tests.

- Evaluation process
  A weightage is associated with each stage of evaluation. This setup can be done by a separate configuration menu accessible by the administrator. This can vary on the basis of the importance of that particular stage in the overall Evaluation process. A sample set up of associated weightage would be -

- Grammar check – 10%
- Spell-check – 10%
- Key phrase matching – 30%
- Abstract – 50%

Hence, in a question carrying 10 marks, 5 marks will be allotted to the abstract matching stage, 3 marks to the key phrase matching stage, while 1 each for the spelling and grammar checks.

- Result Generation process

  This involves the weightage explained above, to arrive at the final marks, along with certain other predefined rules. E.g. if the total count of spelling mistakes is less than a cut off value, say 2, and, if the score is 90% or above in both the key phrase matching and the abstract matching stages, then no mark is deducted for spellings. Another example would be, if the score in the key phrase matching stage is 0, then no marks is awarded even if the other phases have generated some marks as per their individual algorithm. Also, a cut off percentage of accuracy can be entered as an additional parameter in this process, based on which the decision making can be manipulated (e.g. key phrase matching score is not zero but less than this cut off parameter). This setup can also be done by a separate configuration menu accessible by the administrator.

- Student's Quality Evaluation

  This feature is not incorporated in the current model. This is discussed in the section "Future Scope"

VI. CONCLUSION

The model is still in its conceptual stage - initial literature survey, a survey for gathering the seed data for the assessment data bank, and the high level design is complete. Currently coding phase is in progress, after which the system may be tested for internal subjective assessments.

One major challenge is the assessment of answers that are completely essay type and philosophical in nature (mostly associated with the Humanities discipline). To identify key words, key phrases and an abstract for such a scenario would need extensive brainstorming, and consultation with subject matter experts.

VII. FUTURE SCOPE

Student evaluation is complex process which is characterised by human subjectivity. E.g. a student might think in a very original way and completely rephrase what has been taught in the class, while attempting a certain question during the examination. In such cases, there might not be a high percentage of key phrase or abstract match, although the answer is essentially correct. Another aspect of assessment would be a comparative analysis of how the student is improving his understanding of the subject. This is a fundamental requirement of an authentic summative assessment process. This involves introduction of a historical tracking parameter with respect to earlier performances of the particular student in similar tests. Based on this parameter it can be determined, whether the student has performed better or worse, and appropriate advice can be provided. The above two examples show, how the result generation cannot be entirely linear and predefined, as there is an uncertainty element involved. Fuzzy logic can help us deal with such uncertainties. An algorithm based on fuzzy decision making helps to select the optimum model considering a set of criteria and model specifications [11] However in such cases it would not be possible to categorically infer that the student has had zero improvement or maximum improvement. Rather, it would indicate variances that range between the Boolean parameters 0 and 1. As part of the future scope of this research, it is aimed to include the above mentioned feature as part of the "Student's Quality Evaluation" process.

VIII. REFERENCES

[1] Clarke, M. (2012).What matters most for student assessment systems: A framework paper. SABER, Washington D.C.: World Bank.

[2] Cowie, B., & Bell, B. (1999). A model of formave assessment in science education. Assessment in Education: Principles, Policy and Practice, 6, 32-42.
[3] Partnership for Assessment of Readiness for College and Careers (PARCC). (2010, June). The PARCC application for the race to the top comprehensive assessment systems competition. Retrieved October 2, 2010,from www. fldoe.org /parcc /pdf/ apprtcasc .pdf
[4] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[5] Bloxham, S. and Boyd, P. 2007. Developing Effective Assessment in Higher Education. Maidenhead: Open University Press.
[6] Harlen, W. (2006).On the relationship between assessment for formative and summative purposes. In J. Gardner(Eds.), Assessment and Learning. Los Angeles: Sage.
[7] Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18, 119-140
[8] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[9] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[10] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[11] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[12] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[13] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[14] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[15] Black, P. J., &Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7-73.
[16] Earl, Lorna (2003), Assessment as Learning: Using Classroom Assessment to Maximise Student Learning. Thousand Oaks, CA, Corwin Press.
[17] http://www.phpspellcheck.com