

## Automatic Grading and Adaptive Question Selection for English Language Testing

Chitra Bhole<sup>1\*</sup>, Jahanvi Dave<sup>2</sup>, Tanaya Surve<sup>2</sup>, Khushboo Thakkar<sup>2</sup>

<sup>1</sup>Assistant Professor, Department of Computer Engineering, K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India

<sup>2</sup>UG Student, Department of Computer Engineering, K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, Maharashtra, India

\*Corresponding Author: cbhole@somaiya.edu

### ABSTRACT

Computer based evaluation of students' performance is playing a vital role worldwide for ELP examinations. This method is more efficient than manual evaluating process. In this study, a new method is proposed to evaluate the examinees descriptive answers using Text Mining and Text processing algorithms with the help of Natural Language Processing tools. This system includes three sections of ELP, namely: reading, writing and speaking. Dataset for the examination process is extracted using the UiPath platform from reference books which are stored into the database. On taking a test, the system automatically calculates the result and polarity of the subjective answers using NLP and fuzzy logic. In this study, we used Rasch algorithm for adapting to the ability of the examinee for question difficulty level selection and also evaluate the same answer using Natural language processing (NLP) algorithm to check grammar mistakes, polarity and stores the marks in database and finally represent them graphically for better visualization of results.

**Keywords--** Adaptive, automation, english proficiency, test system

### INTRODUCTION

English Language Proficiency (ELP) tests have become a widely recognized way to determine the language skills and abilities of a person. The two most frequently accepted language proficiency tests worldwide are IELTS (International English Language Assessment System) and TOEFL (Test of English as a Foreign Language). Apart from these tests, English tests are also carried out as a part of the placement process for various multinational

companies. Computer based evaluation of student answers is the common work which is used in many areas in assessment of students' learning process. In the learning process, the brilliant concept of using machines has widely changed the world of learning systems. In order to determine the single word answer, for example, decision questions, the computer-based evaluation system was developed and can also determine descriptive responses based on keyword matching. The drawback of this system is that mistakes are impossible to be identified by the candidate which in turn makes it impossible to improve them. Hence, to improve this the proposed new method called evaluating the student descriptive answer using the Natural Language Processing algorithm. Using this raw text can be reduced down to a meaningful set of words and can be checked for their syntactical integrity as well as the polarity of the subjective answers can be drawn.

### LITERATURE SURVEY

Automated Essay Grading System (AEGS) is a system that is based on natural language processing techniques (NLP) and the neural network grading engine is used for automated grading and short essay evaluation, based on a collection of pre-graded essays to determine the student's response and award a score.

An algorithm grading feature based on a neural network that is trained on a set of human pre-graded essays and then measured the similarity between the new essay and the pre-graded essays to allocate a score. writing features analysis consists of a series of programs that analyze and provide input on errors in grammar, usability, and mechanics [1].

The adaptive and automated online assessment is a system proposed by Dumal, P. A. A et al. (2017) is used to evaluate subjective and text-based answers. According to the system, natural

language processing techniques like wordnet, part-of-speech tagging, tokenization and sentence splitting are used for assessing essay type questions. The above system has three main stages. First, searching for the best answers for questions using NLP tools. Second, preprocessing questions and answers by sentence splitting and word tokenization. Third, text analytics which includes keyword analysis and sentence analysis [2].

ASR (Automatic Speech Recognition) technology for English pronunciation correction illustrates the relationship between ELF/ ESL apprentices' self-improvement and English teachers' curriculum teaching. The findings suggest that ASR technology can assist English learners in improving their English pronunciation. The work aims to create a fresh and practical way for ELF/ESL learners to develop English [3].

Meena K and Lawrence Raj conducted a research report on "Evaluation of the Descriptive answers using the Hyperspace Analog to language algorithm and self-organizing Map" The online assessment of the descriptive response based on these works, which would remove the difference in that manual evaluation. To separate categories of terms, the HAL algorithm is used [4].

A Text-to-Speech synthesizer is an application that translates text into spoken word, interprets and processes text using Natural Language Processing (NLP) technology, and then uses Digital Signal Processing (DSP) technology to transform this processed text into a synthesized voice representation. NLP is a research and application approach that explains how to understand, classify, and manipulate a natural language using a programmed (mainly a computer). In the presented document, a text-to-speech synthesizer is used in the form of a simple application that translates input text into synthesized speech and reads it out to the user. It would be of great benefit for visually impaired individuals to build a text to speech synthesizer [5].

## METHODOLOGY

English Language Adaptive Test Series (EPATS): Consider a testing system that integrates the adaptive testing methodology for test generation, Natural Language Processing (NLP) for grading of answers and also makes use of a methodology to assess voice-based answers using software application along with generic computer-based language proficiency test systems. NLP is a technology concerned with the interaction between computers and human language. The field deals with problems ranging from ambiguity resolution both on lexical and syntax level, part-of-speech-tagging (POS), speech and text segmentation, to syntactic and semantic parsing.

### Step 1:

Display the question on a computer-based system and accept the answers of the examinee. Text-based interface can be used that can arrange the answer in the form of tokens by comparing them from an existing file of answers. This can be achieved using Natural Language Processing (NLP) [6].

### Step 2:

Grading of the input answers can be automated using a stack data structure where the corresponding score of each token from the pre-defined file is added subsequently and the final score of a test is marked.

### Step 3:

The marked test will define the next test from the series that the user will have to take. This can be achieved using adaptive methodology. If the result of the previous test was lower than mediocre then, the examinee will be presented with a test of a lower difficulty level and if the performance was higher than mediocre than it will be presented with a test of a higher difficulty level. This will overcome the limitations of a standardized test system where every user has to appear for a standard set of questions which is not the best method to assess an individual's ability (Fig. 1).

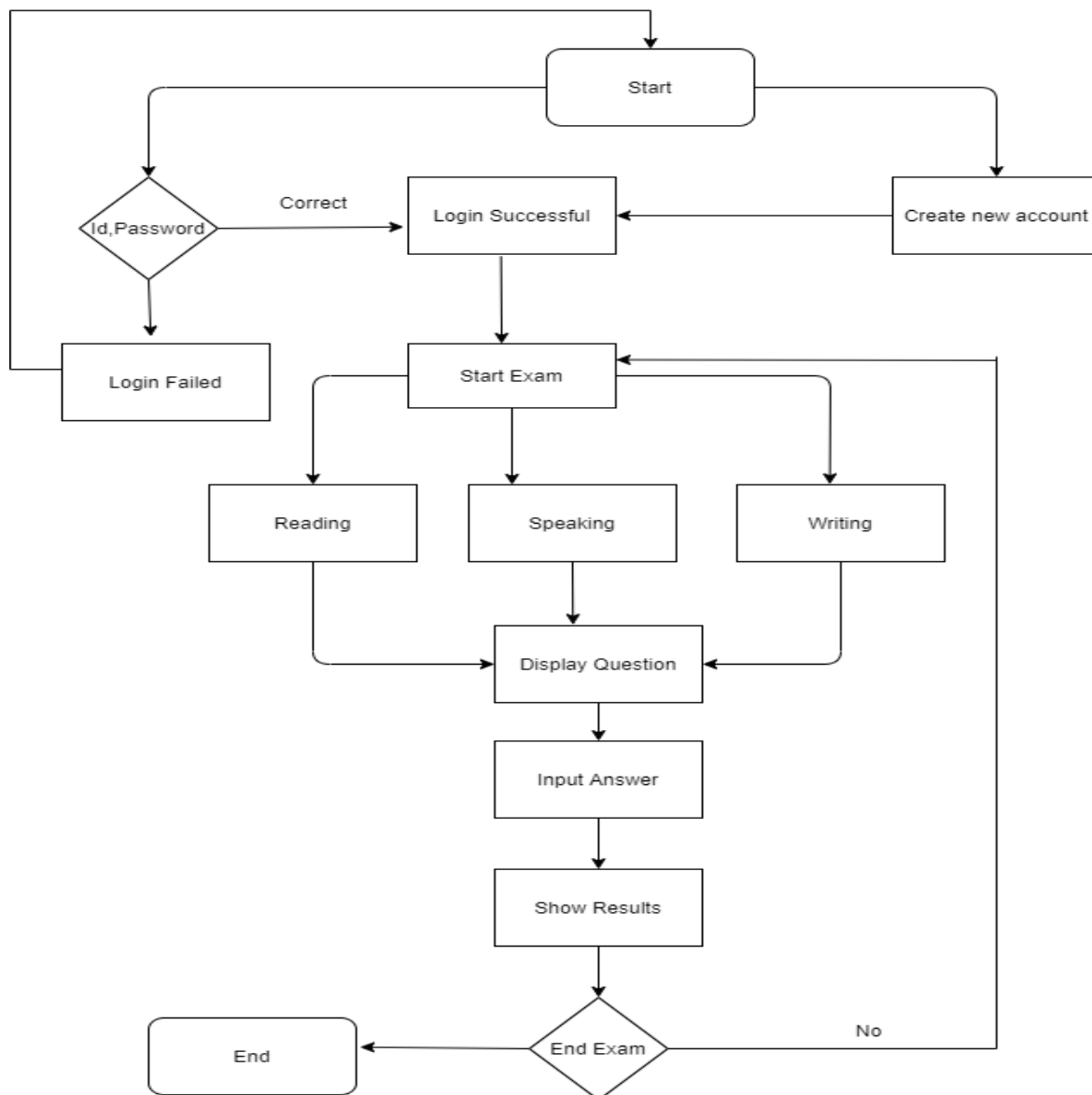


Figure 1: System flow diagram.

### Data Extraction

Traditionally, data in the form of hundreds of records or files are manually extracted from various sources such as books and web. However, this process can be automated such that a question set is extracted from sources without human intervention. UiPath is a Windows desktop (RPA) Software designed to empower business analysts to automate processes. It is a tool based, repetitive task, reducing human power and scalability. It contains less code like .net, scripting language.

### Data Processing

It can be achieved by simple file read operation along with string operation .from a text file of raw data, a single passage can be retrieved by splitting at the occurrence of a certain character. In

a similar way, passage questions can be retrieved by further splitting the passage, this can be viewed as a 2d matrix where each row consists of passage and each cell, a question.

### Check Spelling

The misspelled words are counted in the essay and the performance grade is reduced if they exceed a limit.

### Disable Stopwords

Stopwords are stripped out as these words do not affect the main idea of the essay response such as 'a', 'the', 'is', 'have'.

### Disable Punctuations

Punctuations only contribute to the sentence structure and should not be counted for

analysis; therefore, such characters are removed out of the text before processing.

### Stemming

Stemming is the process by which a written type of words typically reduces corrupted (or often derived) words to their stem, base or root. The stem

does not need to be compatible with the morphological root of the word; it is normally appropriate for similar words to refer to the same stem, although, this stem itself is not a true origin (Fig. 2) [6].

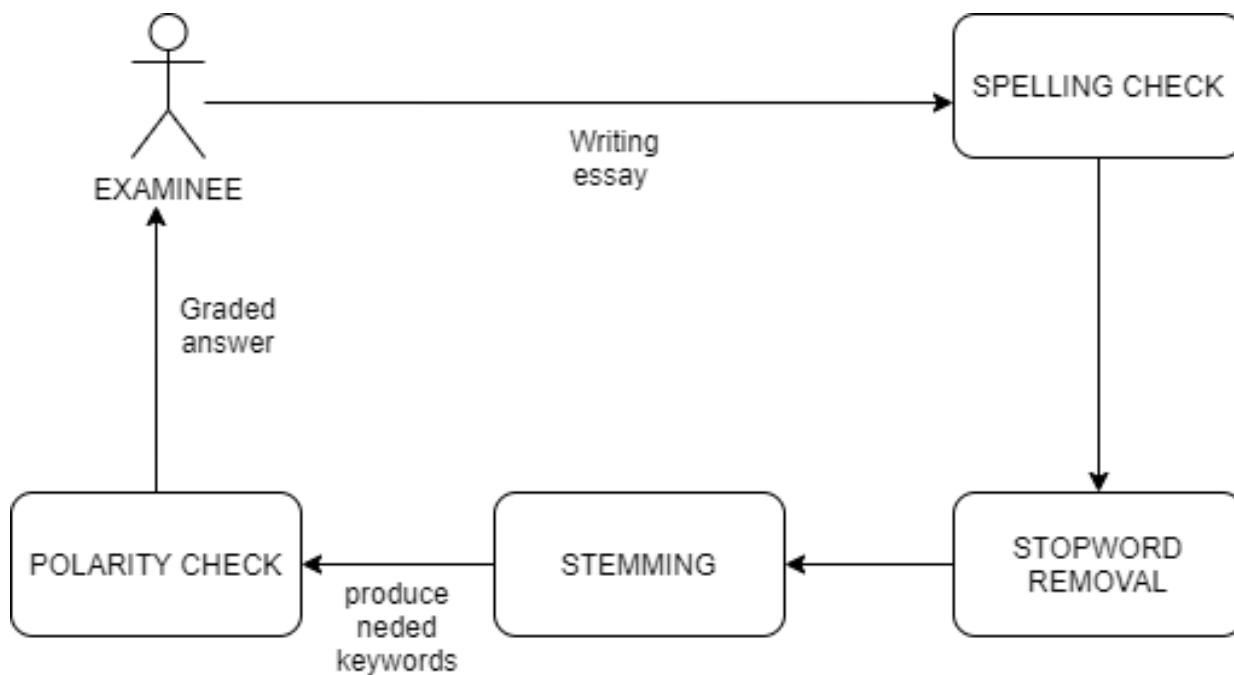


Figure 2: Grading process using natural language processing.

Evaluating Descriptive Type Answers using NLP and getting result:

- Step 1: Select a test from the root question pool.
- Step 2: Input answer in form of speech or text in case of subjective answers.
- Step 3: Check if it satisfies the word limit else mark the word criteria to 0.
- Step 4: Check for syntactical integrity else reduce the syntax parameter for every error by one point.
- Step 5: Check word variety for being below threshold. If less, reduce score for variety.
- Step 6: Score the answer as the average of all three parameters and store it into a database against the user.

Evaluating the students answer using the artificial neural networks algorithm:

- Step 1: Select a test from the root question pool.
- Step 2: Input answer in form of speech or text in case of subjective answers.

Step 3: Remove stop words and stem the remaining words.

Step 4: Using Levenshtein Distance, calculate distance between the actual answer and the user answer.

Step 5: Mark the answer as correct if the match is greater than 65%.

Adaptive question selection algorithm:

ITR (Item Response Theory): Comprises a set of generalized latent variable models designed specifically to respond to data arising dynamically. In the context of the IRT latent variable, this primarily refers to a conceptual concept that describes the observed covariation among the item responses, and secondly to latent responses that are continuous variables underlying observed categorical responses.

The Rasch Model is a type of ITR model that establishes the relationship between a question's complexity and the examinee's ability to respond. It

is based on the underlying logic that subjects have a greater probability of answering easier items correctly and a lower probability of responding to more difficult items [4, 7].

A difficulty group consists of several questions of the same difficulty. All difficulty groups are arranged in ascending order of difficulty (Fig. 3) [8-12].

Step 1: Present a random question from the central difficulty group.

Step 2: Grade the input answer.

Step 3: If the score of the question is greater than 70% then select the next higher difficulty group else select the next lower difficulty group.

Step 4: Remove the answered question from the group.

Step 5: Randomly select the next question from the selected group.

Step 6: Repeat till the end of the test.

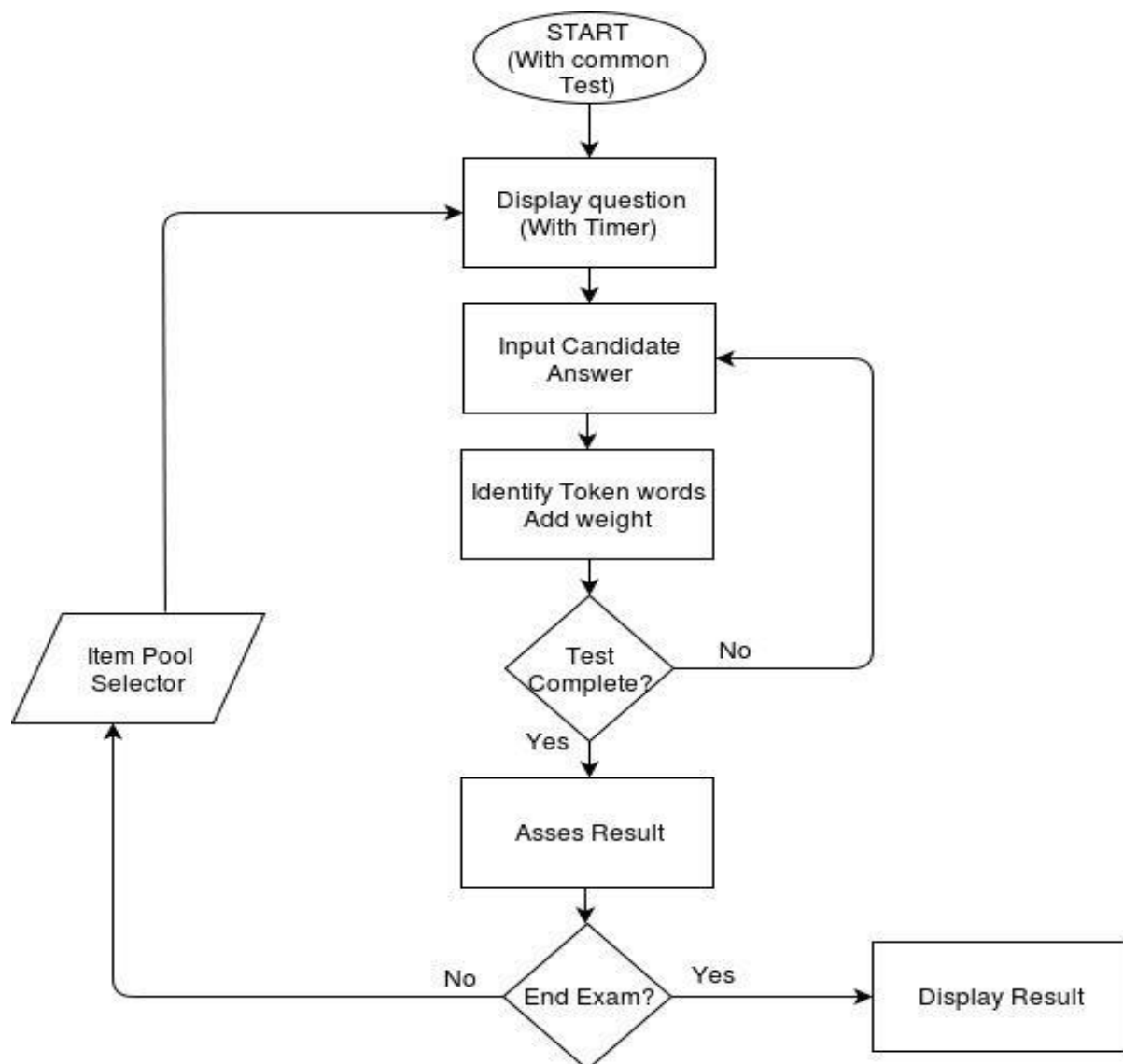


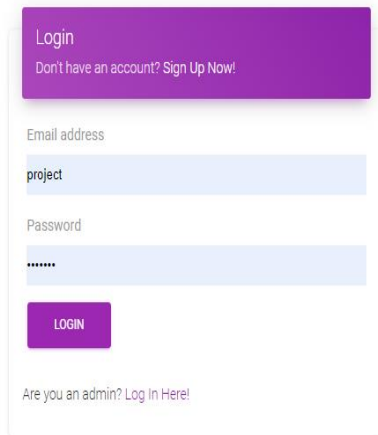
Figure 3: Adaptive test flow diagram.

## RESULTS

The student can take the required test, view their result by logging in to the result page using their respective register number and their password.

There will be 3 different results corresponding to 3 sections:

### Student Login Page



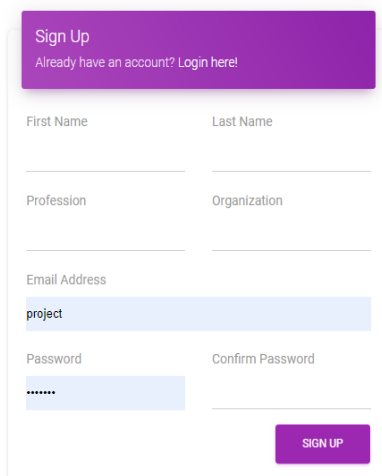
### English Proficiency Adaptive Test Series

**Figure 4: Login.**

This is the student login page, here the student will login to write the examination using the

user id and password created during registration. This provides authentication to the system (Fig. 4).

### Registration Page



### English Proficiency Adaptive Test Series

**Figure 5: Registration.**

This is the registration page of the student, here the student will register their details. This page consists of the name, register number, date of birth, email id,

password. Through these details the student will be authenticated to write the examination (Fig. 5).

## Dashboard

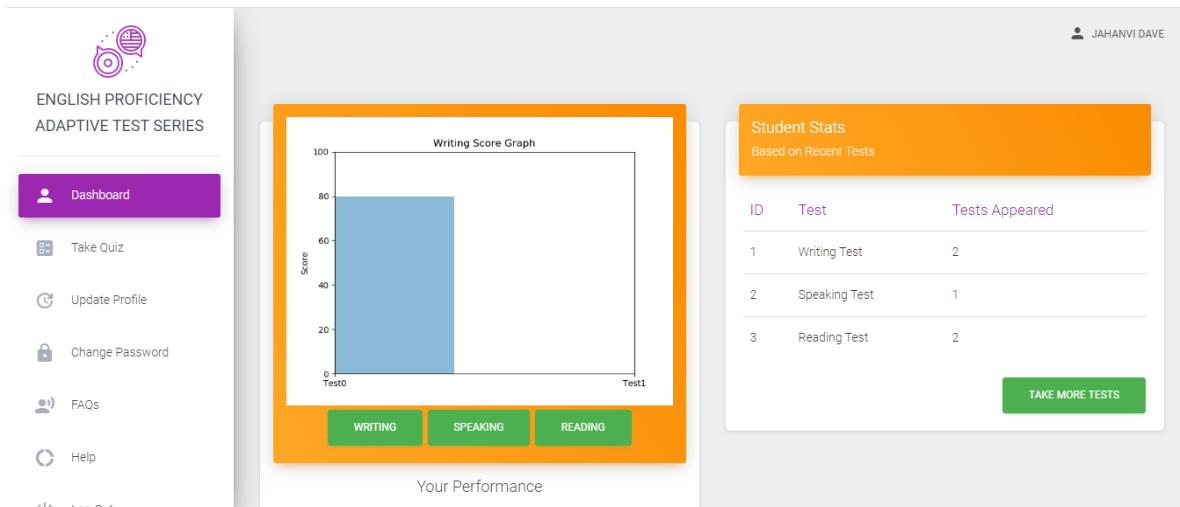


Figure 6: Dashboard.

Keep count of tests taken under each section, display result and progress graph (Fig. 6).

## Take Quiz

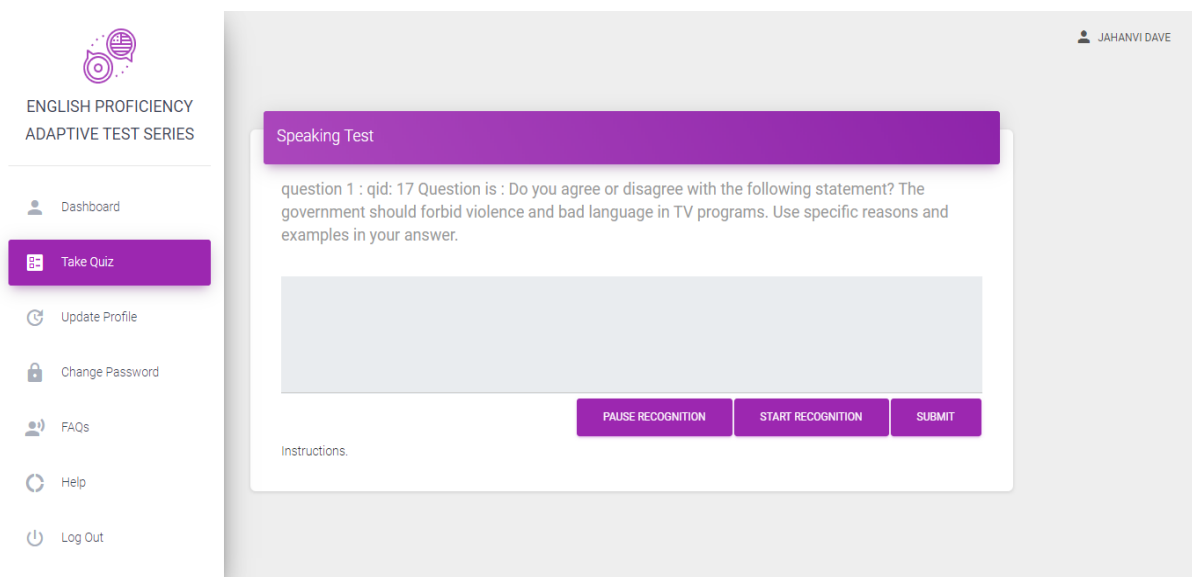


Figure 7: Quiz.

## Individual Tests

Displays questions and provides space for question input (Fig. 7).

## CONCLUSION

EPATS should enable test takers to take an English Proficiency test online. Different types of modules under the test can be attempted within the same platform. The adaptive nature of the system

would help test takers to progress, gradually accelerating the learning process and preparing for competitive tests. Automatic grading of answers provides accurate marking and reduces human effort. Performance prediction parameters can be calculated and improvement scope can be defined for each examinee individually.



#### ACKNOWLEDGEMENT

We would like to convey our sincere thanks to everyone who guided us throughout our research. We wish to express our sincere thanks to Dr. Suresh Ukarande, our Principal, for providing us with all the necessary facilities for the research. We place on record, our sincere thank you to Prof. Sarita Ambadekar, Head of Department, for the continuous encouragement.

#### REFERENCES

1. Ghosh S., Fatima S. S. (19-21 November, 2008), "Design of an automated essay grading (AEG) system in the Indian context" *TENCON 2008 - 2008 IEEE Region 10 Conference*, Hyderabad, India.
2. Dimal, P. A. A., Shanika, W. K. D., Pathinayake, S. A. D., Sandanayake, T. C. (6-8 December, 2017), "Adaptive and automated online assessment evaluation system", *11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, Malabe, Sri Lanka.
3. Li M., Han M., Chen Z., Mo Y., Chen X., Liu X. (27-29 June, 2017), "Improving english pronunciation via automatic speech recognition technology", *International Symposium on Educational Technology (ISET)*. Hong Kong, China.
4. Ismail A., Roslan L., Adnan A. N. (9-10 November, 2017), "Assessment on course outcome performance using the Rasch measurement model", *IEEE 9th International Conference on Engineering Education (ICEED)*, Kanazawa, Japan.
5. Mukherjee P. M., Santra S., Bhowmick S., Paul A., Chatterjee P., Deyasi A. (2018), "Development of GUI for text-to-speech recognition using natural language processing", *2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)*.
6. Shehab A., Elhoseny M., Hassanien A. E. (28-29 December, 2016), "A hybrid scheme for automated essay grading based on LVQ and NLP techniques", *12th International Computer Engineering Conference (ICENCO)*, Cairo, Egypt.
7. Bond TG, Fox CM (2001), "Applying the Rasch model: fundamental measurement in the human sciences", Mahwah: Lawrence Erlbaum Associates.
8. Guo Guiran, Zhao Shul (9-11 July, 2010), "Research on the computer-based college English test system in China", *3rd International Conference on Computer Science and Information Technology*, Chengdu, China.
9. Ahkour K., Machkour M., Ennaji M., et al. (22-24 July, 2019), "Comparative study of existing approaches on the task of natural language to database language", *International Conference of Computer Science and Renewable Energies (ICCSRE)*, Agadir, Morocco.
10. Wang Y., Sun Y., Chen Y. (16-19 August, 2019), "Design and research of intelligent tutor system based on natural language processing", *2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI)*, Kunming, China.
11. Meena K, Lawrence Raj (18-20 December, 2014), "Evaluation of the descriptive answers using the hyperspace analog to language and self-organizing map", *IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, India.
12. Shengnan Zhang, Yan Hu, Guangrong Bian (25-26 March, 2017), "Research on string similarity algorithm based on Levenshtein distance", *IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China.