Ministry of Higher Education & Scientific Research

University of AL-Qadisiyah

College of Computer Science & Information Technology

Department of Computer

Essay question Assessment in E-learning

Ander Graduating Project

Mohammed Hassan Khudair

Hussein Mohammed Jawad

Walaa Mohammed olaiwy

Baneen Hussein Hameed

College of Computer Science & Information Technology

University of AL-Qadisiyah

Under the supervision of

Assistant lecturer

Manar Joundy Hazar

بسم الله الرحمن الرحيم

﴿ وَقُلِ اعْمَلُوا فَسَيَرَى اللَّهُ عَمَلَكُمْ وَرَسُولُهُ وَالْمُؤْمِنُونَ ﴾

[سورة التوبة: (١٠٥)]

الاهـــــــــــــداء

بدانا بأكثر من يد وقاسينا أكثر من هم وعانينا الكثير من الصعوبات وها نحن اليوم والحمد لله نطوي سهر الليالي وتعب الإيام وخلاصة مشوارنا بين دفتي هذا العمل المتواضع. إلى منارة العلم والامام المصطفي إلى الأمي الذي علم التعلمين إلى سيد الخلق إلى رسولنا الكريم سيدنا محمد ﷺ.

إلى الينبوع الذي لا يمل العطاء إلى من حاكت سعادتي بخيوط منسوجة من قلبها إلى والدتي العزيزة. إلى من سعى وشقى لأنعم بالراحة والهناء الذي لم يبخل بشيء من أجل دفعي في طريق النجاح الذي علمني أن أرتقي سلم الحياة بحكمة وصبر إلى والدي العزيز.

إلى من حبهم يجري في عروقي ويلهج بذكراهم فؤادي إلى أخواتي وأخواني. إلى من سرنا سوياً ونحن نشق الطريق معاَ نحو النجاح والإبداع إلى من تكا تفنا يداً بيد ونحن نقطف زهرة وتعلمنا إلى صديقاتي وزميلاتي.

إلى من علمونا حروفا من ذهب وكلمات من درر وعبارات من أسمى وأجلى عبارات في العلم إلى من صاغوا لنا علمهم حروفا ومن فكرهم منارة تنير لنا سيرة العلم والنجاح إلى أساتذتنا الكرام.

Contents

Abstract:

E-Learning is the employment of technology to support in improving learning manner. It is considered as an integral part of instruction-based learning methods. Assessment is a method of evaluating student knowledge based on specific learning objectives. Intelligent Tutoring System (ITS) provides an singular tutoring system and it is considered as a virtual training assistant to develop student's knowledge without instructors. A framework is proposed in this paper to evaluate students essay answers based on a linguistic knowledge. The gotten results from implementing our proposed system in evaluating our simulation results indicated a high accuracy in its performance when compared with the other techniques. The efficiency of this framework was measured using Pearson's correlation coefficient and Root Mean Square Error (RMSE) based on a student answers evaluation obtained from North Texas University dataset, which is available online.

Keywords:

Intelligent Tutoring System, Word Net, Natural Language Processing, Semantic Similarity, Student Assessment, Summarization.

Chapter One

INTRODUCTION

1.1 INTRODUCTION

E-learning is an effective way of teaching using the Internet. With e-learning, you can provide courses for students to study at any time and from anywhere, as well as interact with them in an easy and effective way. It is clear that e-learning has become one of the requirements of the educational process not only to keep abreast with the current developments and very quickly in educational institutions around the world, but also for e-learning to play a real role in improving education and its outcomes [1].

In addition to E-Learning can be identified as the deploying technology to help and to develop learning. It starts with televisions, projectors then computer programs, real-time online discussion groups, video and telephone conferencing and 3D simulations [2], [3]. Intelligent Tutoring System (ITS) is used to provide individualized tutoring or instruction. It used to as training assistant in a specific subject and teaching expertise's providing captivating new options. It includes instructional model to clarify exactly teaching subjects (what should teach) teaching strategies (How to teach).

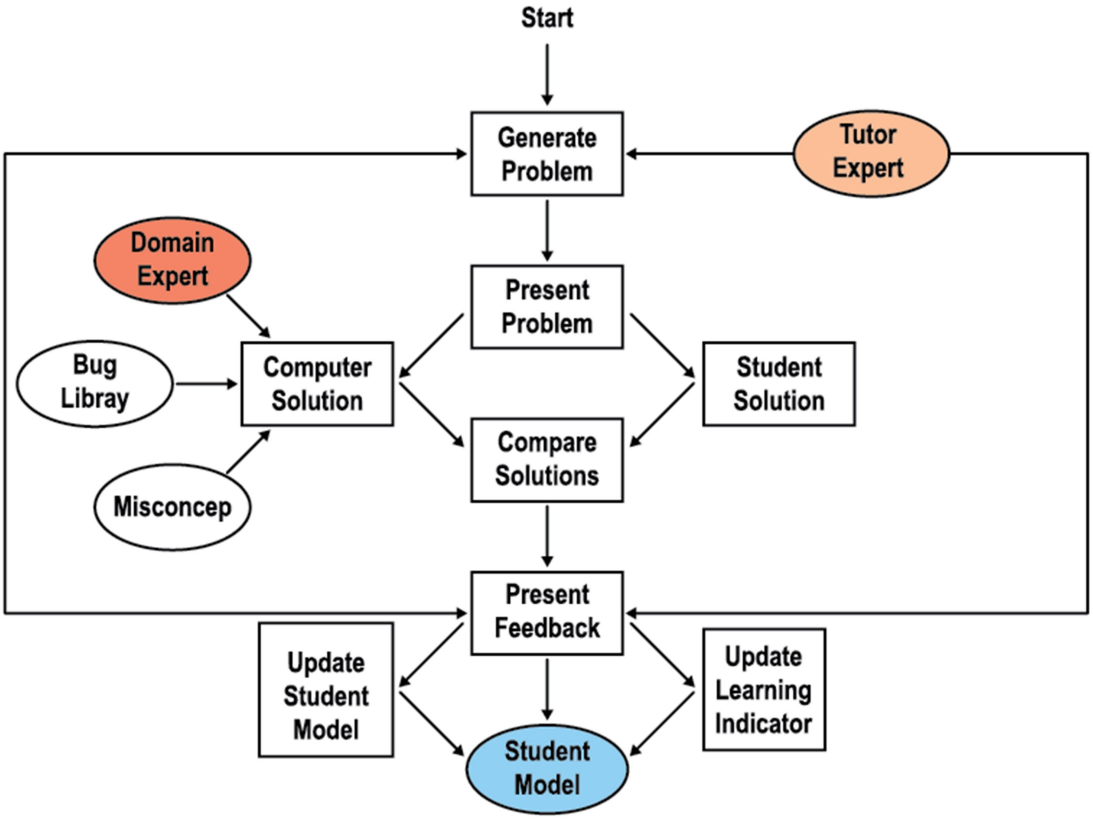

Fig. (1) Intelligent Tutoring System Main Components

The assessment of learning outcomes with tests and examinations can be facilitated by many types and grading methods. The specific question types may be designed as anything from simple multiple-choice questions, to questions requiring natural language responses such as short answers or essays. The grading method may be either manual grading by hand or automatic grading by computational methods. In this paper we focus on the short answer question type and the automatic grading method. We refer to this field as automatic short answer grading, or ASAG. Assessment is considered to play a central role in the educational process. (Valenti et al., 2003).

One of the difficulties of grading essays is the subjectivity, or at least the perceived subjectivity, of the grading process. Many researchers claim that the subjective nature of essay assessment leads to variation in grades awarded by different human assessors, which is perceived by students as a great source of unfairness.

This issue may be faced through the adoption of automated assessment tools for essays. A system for automated assessment would at least be consistent in the way it scores essays, and enormous cost and time savings could be achieved if the system can be shown to grade essays within the range of those awarded by human assessor. Furthermore, according to Hearst (2000) using computers to increase our understanding of the textual features and cognitive skills involved in the creation and in the comprehension of written texts, will provide a number of benefits to the educational community. In fact, “it will help us develop more effective instructional materials for improving reading, writing and other communication abilities. It will also help us develop more effective technologies such as search engines and question answering systems for providing universal access to electronic information [5].

Apex (for an Assistant for Preparing Exams), a tool for evaluating student essays based on their content. It relies on a semantic text analysis method called. Latent Semantic Analysis (LSA) [7]. Basically, LSA represents

each word of a text as a vector in a high-dimensional space such that the proximity between two vectors is closely related to the semantic similarity between the two corresponding words. Since a text is composed of words, this measure of similarity can be easily extended to compare texts. Thus, it is possible to compare two student essays, or a student essay and the text of a course, on a semantic basis.

Many college and university professors desire to change their instructional style from traditional lecture to a more active, student-centered style through the use of group projects, discovery activities, experiments, and class presentations (Baker, 2000). Until recently, however, some professors feared they would sacrifice course content if they utilized such active learning techniques during class. With the increased availability of web-based instructional technologies like Web CT and Blackboard during the late 1990s, professors began providing students with access to course content via video and PowerPoint lectures outside the classroom (Lage & Platt, 2000). With the delivery of course content secured via technology, professors felt freer to introduce activities inside the classroom that would give students the opportunity to engage material in an environment where other students and the professor are present to aid in the learning process.

This change in how course content is introduced to and engaged by students is a significant departure from the lecture-homework cycle found in more traditional classrooms. Perhaps the most striking departure is the physical location of where the introduction and deeper engagement with the material occurs. Traditionally, the introduction is given in class through a lecture, and the deeper engagement occurs outside of class through homework. In the above description, however, the introduction occurs outside of class and the engagement occurs inside the classroom [9].

In this field most researchers agree on that some aspects of complex achievement are difficult to measure using objective-type questions. Learning

outcomes implying the ability to recall, organize and integrate ideas, the ability to express oneself in writing and the ability to supply merely than identify interpretation and application of data, require less structuring of response than that imposed by objective test items. It is in the measurement of such outcomes, corresponding to the higher levels of the Bloom's (1956) taxonomy (namely evaluation and synthesis) that the essay question serves it's most useful purpose [10].

1.2 PROBLEM STATEMENT

Not the same of Multiple Choice Question (MCQ), essays enclose subjective answers rather than the accurate answers in MCQ (*e.g. true or false*) (Kakkonen & Sutinen 2004). So, the student skill and ability play a large role in creating a strong answer free of misspellings and grammatical errors that will reduce the student's mark if they exist. Therefore, the process of automated essays evaluations is a challenging task because of the need of comprehensive evaluation in order to validate the answers accurately (Landauer 2003).

The difference between say multiple choice and short answer questions is easy to comprehend, but the difference between other question types such as short answers and essays can become blurred. Therefore, we say that a short answer question is one that can be considered as meeting at least five specific criteria. First, the question must require a response that recalls external knowledge instead of requiring the answer to be recognized from within the question. Second, the question must require a response given in natural language. Third, the answer length should be roughly between one phrase and one paragraph. Fourth, the assessment of the responses should focus on the content instead of writing style. Fifth, the level of openness in open-ended versus close-ended responses should be restricted with an objective question design [13].

A framework is proposed in this project to evaluate students essay answers based on a linguistic knowledge.

1.3 IMPORTANCE OF STUDY

Much work has been conducted in the field of automatic grading but the systems are mainly based on multiple choice exams [14]. These grading programs are not hard to make. The difficulty lies rather in the design of the propositions which should be close enough to the right answer but still wrong. An alternative to multiple choice tests is to ask the student to write an essay about what he or she knows about a domain and then to compare that text to pre-graded texts.

Educators in Britain are spending about 30% of their time in assessing and grading student’s answers, which produces a loss of an expected 3 billion pounds per year. Therefore, it can be imagined that much benefits such as improving economy and saving time, are gained from the application of automated essays grading systems. Automated assessment of students' free-text answers has several challenges [15].

Chapter Two

Related Work

2.1 Research Background

Automatic Essay Scoring (AES) is the study that has been proposed to assess the teachers by providing an automatic approach to evaluate the score of an essay. In fact, there are several techniques have been used for AES where the writing style, lexical analysis, semantic analysis, syntactic analysis and probabilistic approach have been examined in terms of providing scores [16].

2.1.1 Editorial questions:

This is the oldest pattern of common test questions used since ancient times. These questions allow the student to answer the question in the form of an essay that is formulated in his or her own style and usually requires expressive or structural answers, giving the student an opportunity to express his ideas using his ability to the creation of interrelated sentences, and such questions have certain criteria that must be adhered to when they are written in the exams. Here the student skill and ability play a large role in creating a strong answer free of misspellings and grammatical errors that will reduce the student's mark if they exist [17].

2.1.2 Criteria for writing essay questions:

- To consider the areas in which these questions are available, such as: limited number of students, or limited to the educational outputs of the higher grades, so it must be depending on the situation, the purpose and the goal.
- Taking into account the development of a good plan during preparation, adherence to procedures and steps to prepare them. Careful to choose the appropriate and clear words and forms for the question, where the student understands it, and examples of these formulas: (Discuss, explain, compare), and therefore can answer correctly.
- Keep away from missing formulas, and open during the development of questions.

- Consistency between the required achievement and the nature of the questions.
- The words used should be used as a function of the category and quality of the question, such as: (compare, in terms of, or deny) the avoidance of the use of formulas for substantive questions, such as where, what, and when. Keep the questions covered in all content, and the goal that students are expected to achieve, as they take into account more essay questions, with attention to the time required for each answer.
- Define the typical answer to the questions, which are adopted during the correction, taking into account the most important elements to be mentioned as a complete question mark. Also, be careful not to neglect the sub-answers, which reduces and restricts the teacher's ability to choose the answer he wishes and to see it correctly according to his mood [18].

2.1.3 Advantages of essay questions:

1. Allows students to select the right facts and ideas, and is primarily free to choose and organize answers.
2. To suit all the abilities and possibilities of the student, helping him to link his ideas, information, and briefings to be integrated and adequate.
3. Allows students to discover their abilities in finding solutions to problems, by employing the correct knowledge.
4. This allows the student to express in the way he wants, which contributes to the discovery of his culture and knowledge, as well as a fertile field to reveal his terminology and information and confirm its validity, thus providing the teacher with the ability to evaluate it based on its stock, abilities and skill in expression.
5. The proportion of these questions is small compared to the objective questions [17,18].

2.1.4 Disadvantages of essay questions:

1. The teacher does not allow the teacher to inform the entire curriculum due to the lack of questions because of the long time it takes to answer. The teacher cannot put many questions in the exam to take into consideration the abilities of the students and their ability to complete the exam within the specified time. Standards for all educational outputs.
2. Do not give the student the right to a satisfactory mark, if the teacher has corrected randomly, according to his temperament, and without taking into account the typical answers, or because he sees the correct answer is different from what the student answered, especially as such questions provide the teacher with the opportunity to intervene in the answer and determine.
3. It does not take into account the accuracy of marking, which is necessary and necessary in the development of questions [17,18].

2.2 Related work

several researchers have addressed the problem of automated essay scoring or so-called automatic essay assessment by using various techniques. The key characteristic behind these techniques lies on a set of manually scored essay by human in which the essay that intended to be assessed is compared with the pre-scored essays. Usually, the manually scored essays are called pre-scored essays or training essays, whereas the essay that required to be assessed by the computer is called tested essay or automated scoring essay [19].

The earliest system proposed for essay assessment is the Project Essay Grader (PEG) (Page 1966) which was focusing on the writing style of a given essay in order to provide the score. The writing style concentrates on essay length and mechanics such as spelling error, capitalization, grammar and diction. Obviously this approach was criticized due to the lack of semantic analysis in which the content is being ignored.

A question generation system was presented by Yao et al. based on the approach of semantic rewriting. State-of-the-art deep linguistic parsing and generation tools are working to map natural language sentences into their meaning representations in the form of Minimal Recursion Semantics (MRS) and vice versa. a principled way of generating questions is obtained, which avoids the ad-hoc manipulation of syntactic structures. Based on the (partial) understanding of the sentence meaning, the system creates questions that are semantically grounded and purposeful.

There are many methods that used in Graesser, Arthur C., et al. (2000), which is correcting the computer for the article presented in the test answers by sample of students and one of the methods used to evaluate the article provided by a group or sample of students using the model of efficiency based on educational education. One scientist suggested a model or approach to lessons the smart test answers students using a grammar checker and providing feedback after evaluating the student's answer. Another world - class student identification software program proposes using a check-in-side flow control diagram to measure the similarities. Others used scoring algorithms to measure students' writing articles, their overall knowledge, comprehension, and reading skills. The purpose of their work is to link measurements to the mentioned standards with a smart writing teacher. The results of his experiments showed that the scoring algorithm applied was less in line with both the comprehension and cognitive measurements of the human evaluations... and examined the impact of applying hierarchical technique in rating article evaluations.

Satav, Harshada, Trupti Nanekar, Supriya Pingale, and Nupur [35] has presented an examination system that based on SQL and Microsoft.Net (C# and Asp.Net). They implemented an examination system for computer application base. Their system has only examination and evaluation subsystems. Moreover, their system included different types of questions that

related to computer application such as multiple choice, fill in blank, true / false, programming design questions.

Ge Yu, Libin Hong, and Lei Sheng [34] has developed a web-based examination system and evaluation system for computer programming. Their system has an examination and exercising systems for programming. Furthermore, it includes preposition subsystem for teachers in order to manage the set of exercises and questions. It also includes a monitoring system which is used to configure the exam settings. In addition, they provided an examination system that provides the examination task. They used a fill in blank evaluation method that is based on the separation of the key words and matching them with the answers key. Mohamed Jaballah, Saad Harous, and Sane M. Yagi [36] developed an Arabic examination system for students in University of Sharjah. Their system just included examination and grading subsystems of different types of questions such as true/false, multiple choice, fill in blank and essay question types. The exam paper is generated automatically by the examination system. However, their grading subsystem was not automated and the grading is done manually by the teacher via a grading portal.

Chen Xiangjun and Wu Fangsheng [37] proposed an examination system which provides login activity recording, users management, test question management. It includes examination subsystem and grading subsystem that based on matching the student answers with the answers key.

On the other hand, our WBSECIL includes examination subsystem, smart grading subsystem, homework submission subsystem, smart discussion board subsystem, and administration subsystems. In addition, it employs AI algorithms to smartly grade the fill in blank questions by measuring the sentences similarity.
Syntax-based Approach. In this approach, processing follows a common

strategy for any input sentence. This strategy is summarized as four basic steps as :

1. Parsing the sentence to determine the syntactic structure: Sentence detection is the first and most major pre-processing step in the question generation process. Sentence detector is the main component of any processing framework relevant to natural language that concerns splitting input text whatever a whole document, paragraph or sentence.
2. Simplifying the sentence if possible: Sentence simplification is necessary because it makes some aspects of question generation easier. This process uses one or more simplification steps, including splitting sentences containing independent clauses, appositive removal, prepositional phrase removal, discourse marker removal, and relative clause removal. While simplification makes some aspects of question generation easier, it also introduces new problems that must be handled, such as level of simplification required (separately or in a combined mode), and processing different types of clauses (e.g. illative, concessive, conditional, consecutive, adjectival, or adverbial

2.3 AUTOMATIC ESSAY SCORING APPROACHES

In order to understand the mechanism of automated scoring systems, the approaches that have been used by the previous researches should be illustrated in details. As mentioned earlier, the automatic essay scoring depends mainly on manual pre-scored essays by human in order to be compared with the new tested essays. In particular, the mechanism of such comparison is conducted using several approaches. One of the earliest approaches is the writing style in which the pre-scored essays are compared with the new tested essay in terms of number of paragraphs, number of sentences and number of words (Alghamdi et al. 2014; Burstein et al. 2003; Page 2003). This can be conducted by identifying pre-scored essay that share the same writing style characteristics of the tested essay. In this manner, the

score of the most similar pre-scored essay will be assigned to the new tested essay. For example, if the new tested essay contains five paragraphs, its automatic score will be the same with the pre-scored essay that contain five paragraphs.

However, other approaches aim to conduct the comparison between the pre-scored essays and the new tested essay based on the content of these essays (Gomaa & Fahmy 2014). In this manner, lexical analysis could be used in order to examined lexical similarity between their words. For example, if a pre-scored essay contains particular word such as ‘plant’ and the tested essay contains the same word but with derivation such as ‘planting’, lexical analysis has the ability to identify such similarity. In addition, lexical analysis is useful approach to identify the most frequent terms of the two essays. In this case, it is easy to address the similarity between the frequent terms from the two essays.

Other approach would be used in the essay scoring is the semantic analysis in which the similarity between two essay could be conducted based on the meaning of words such as the ‘plant’ and ‘grass’ (Refaat et al. 2012). This capability is not provided by lexical analysis. However, to apply the semantic analysis, an external knowledge resource has to be provided such as dictionary, thesaurus or lexicon. There are some available dictionaries such as WordNet but it is associated only with English language. This can put a limitation for other languages such as Arabic. Therefore, some researchers come up with new semantic methods that do not require the use of dictionary. These methods are mainly depending on statistics such as Latent Semantic Analysis (LSA) and Distributional Semantic Co-occurrence (DISCO).

Other researchers have focused on the syntactical or grammatical analysis in which the verbs, nouns, adjectives are being analyzed with their semantic (Kanejiya et al. 2003). In addition, noun phrases and verb phrases

are being divided in order to establish an independent comparison between the pre-scored essay and the new tested essay.

C O N C L U S I O N

This project describes a system that can assess student essays based on their content from different points of view. In Our experiment provide promising results. In particular, we show a significant interesting correlation between human grades and Apex grades. This can lead to the development of automatic grading systems not only based on multiple choice exams or surface features of essays, but rather on semantic features of unrestricted essays. This approach can also be used in a distance learning context since students can connect to the system and freely submit essays. The system evaluates the essays as many times as required by the students, without getting bored. Moreover, there is no need for the teacher to code any domain knowledge. The text of the course is all that is required.

References

1. Motiwalla, L. F. (2007). Mobile learning: A framework and evaluation. *Computers & education*, *49*(3), 581-596
2. S. Sarhan, *"Intelligent Tutoring System", M.Sc. Thesis, Dept. of Computer Science, Faculty of Computers and Information, Mansoura University,* 2009
3. S. Sarhan, R. Bahgat, A. Tolba, "Rough-Neuro Model for Improving Student State Diagnosis in Intelligent Tutoring System", *Egyptian Rough Computing Journal (ERCJ),* 2009.
4. Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education: Research*, *2*, 319-330.
5. Bin, L., & Jian-Min, Y. (2011, September). Automated essay scoring using multi-classifier fusion. In *International Conference on Information and Management Engineering* (pp. 151-157). Springer, Berlin, Heidelberg.
6. Hou, W. J., & Tsao, J. H. (2011). AUTOMATIC ASSESSMENT OF STUDENTS'FREE-TEXT ANSWERS WITH DIFFERENT LEVELS. *International Journal on Artificial Intelligence Tools*, *20*(02), 327-347.
7. Felder, R. M., Woods, D. R., Stice, J. E., & Rugarcia, A. (2000). The future of engineering education II. Teaching methods that work. *Chemical Engineering Education*, *34*(1), 26-39.
8. Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*, *31*(1), 30-43.

9. Strayer, J. (2007). *The effects of the classroom flip on the learning environment: A comparison of learning activity in a traditional classroom and a flip classroom that used an intelligent tutoring system* (Doctoral dissertation, The Ohio State University).
10. Ghosh, S. (2010). Online Automated Essay Grading System as a Web Based Learning (WBL) Tool in Engineering Education. *Web-Based Engineering Education: Critical Design and Effective Tools: Critical Design and Effective Tools*, 53.
11. Kakkonen, T., & Sutinen, E. (2004, June). Automatic assessment of the content of essays based on course materials. In *Information Technology: Research and Education, 2004. ITRE 2004. 2nd International Conference on* (pp. 126-130). IEEE.
12. Landauer, T. K. (2003). Automatic essay assessment. *Assessment in education: Principles, policy & practice*, *10*(3), 295-308.
13. Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, *25*(1), 60-117.
14. Lemaire, B., & Dessus, P. (2001). A system to assess the semantic content of student essays. *Journal of Educational Computing Research*, *24*(3), 305-320.
15. Lenz, B., Wells, J., & Kingston, S. (2015). *Transforming schools using project-based deeper learning, performance assessment, and common core standards*. John Wiley & Sons.
16. Islam, M. M., & Hoque, A. L. (2010, December). Automated essay scoring using generalized latent semantic analysis. In *Computer and Information Technology (ICCIT), 2010 13th International Conference on* (pp. 358-363). IEEE.
17. Heywood, J. (2000). *Assessment in higher education: Student learning, teaching, programmes and institutions* (Vol. 56). Jessica Kingsley Publishers.
18. Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. University of Chicago Press.

19. Hermet, M., & Szpakowicz, S. (2006). Symbolic assessment of free text answers in a second-language tutoring system.