A Framework for Automatic Exam Generation based on Intended Learning Outcomes

Ashraf Amria[1], Ahmed Ewais[2,3] and Rami Hodrob[2]
[1]Deanship of Admission, Registration & Examinations, Al- Quds Open University, Jenin, Palestine
[2]Department of Computer Science, Arab American University, Jenin, Palestine
[3]Web & Information Systems Engineering (WISE) Laboratory, Vrije Universiteit Brussel, Brussels, Belgium

Keywords: Bloom's Taxonomy, Learning Outcomes, Automatic Exam Generator, Assessment.

Abstract: Assessment plays important role in learning process in higher education institutions. However, poorly designed exams can fail to achieve the intended learning outcomes of a specific course, which can also have a bad impact on the programs and educational institutes. One of the possible solutions is to standardize the exams based on educational taxonomies. However, this is not an easy process for educators. With the recent technologies, the assessment approaches have been improved by automatically generating exams based on educational taxonomies. This paper presents a framework that allow educators to map questions to intended learning outcomes based on Bloom's taxonomy. Furthermore, it elaborates on the principles and requirements for generating exams automatically. It also report on a prototype implementation of an authoring tool for generating exams to evaluate the achievements of intended learning outcomes.

1 INTRODUCTION

One of the means to measure the impact and the output of learning process in schools is the use of assessment techniques. In general, assessment plays an important role in supporting the learning process of the students. This support is achieved by evaluating the students' results and answers using some automatic tools. This will provide stakeholders good vision and overview of the learning process.

Recently, the era of education is complemented by effective utilization of technology For instance, developing learning materials using different applications, and using Virtual Reality and Augmented Reality is used in many different domains. Furthermore, distance learning and e-learning are also good examples of the use of recent technology. In many domains, learners can get certificates from higher education institution using Massive Open Online Courses (MOOCs) without being limited to the place and time. Providing certificates can be based on evaluating student's achievements after following the online course. Therefore, electronic exams have been used in a wide range of domains to measure the effectiveness of learning process. For this purpose, researchers proposed different approaches for generating exams to evaluate the effectiveness of the learning process (Manuel Azevedo et al., 2017).

One way to guarantee a correct measurement of the intended learning outcome (ILO) of a specific course module in higher education institutions is to provide proper questions that effectively measure the intended learning outcome in the conducted exams, exercises, quizzes, etc. An approach for realizing such an effective assessment tools is to relate learning outcomes with both learning topics and questions related to each learning topic. For instance, in (Blumberg, 2009) an approach for maximizing the learning process by aligning learning objectives, learning materials, activities, and course assessment with Blooms' taxonomy is proposed. The alignment is done using action verbs of the different levels of cognitive process. Other researchers (Tofade et al., 2013) proposed some best practices for using questions in course modules. Among the proposed practices of using educational taxonomies is that of the use of Bloom's taxonomy to define different levels of questions.

In general, an educational taxonomy is used to describe the learning outcomes using the courses syllabi. Furthermore, educational taxonomies can be used to provide an overview about the different level of understanding about specific learning concepts and topics. Another important aspect for the use of

Amria, A., Ewais, A. and Hodrob, R.
A Framework for Automatic Exam Generation based on Intended Learning Outcomes.
In *Proceedings of the 10th International Conference on Computer Supported Education (CSEDU 2018) - Volume 1*, pages 474-480
ISBN: 978-989-758-291-2

educational taxonomies in the learning process is to identify the level of exams' questions depending on cognitive levels. For instance, course exams should include questions that asses different level of learning effectively.

Based on educational and pedagogical theories, researchers proposed different taxonomies to help educators in developing learning resources, assessments, and learning outcomes. Among the proposed taxonomies is the Bloom's taxonomy (Bloom, 1956) and its revised version (Krathwohl, 2002). It is mainly based on six levels of the cognitive learning process: Remember, Understand, Apply, Analyze, Evaluate, and Create. Furthermore, a list of different action verbs has been identified to describe the intended learning outcomes of a course. The revised version of the Bloom's taxonomy is mainly mapping cognitive dimensions with the knowledge dimensions. Another taxonomy is the so-called SOLO taxonomy (Biggs & Collis, 1982) which has the levels: Prestructural, Unistructural, Multistructural, Relational, and Extended Abstract, and which are not only restricted to cognitive aspects but also deal with knowledge and skills. More educational taxonomies that are used in the assessment and evaluation are reviewed in (Fuller et al., 2007; Ala-Mutka, 2005).

In general, there are three types of exam generation approaches (Cen et al., 2010). The first type is related to offering a question repository that can be explored by educators to select the questions for a specific exam. This type is almost similar to the manual creation of the exam. However, the educators can inspect the stored questions in the database by means of a user interface. The second type is related to generating the exam based on random selection of the questions. The third type is related to generating the exams by means of AI algorithms for realizing predefined rules to provide the exam.

Normally, identifying simple or difficult questions is mainly depending on the educators' intuition and experience. Furthermore, similar questions or repetition of questions can happen in manual created exams. Another possible drawback is related to careless division of the total mark of the exam over the composed questions. Finally, manual preparation of exams with the alignment of questions and learning outcomes requires a high mental demand. Given the previous drawbacks, there is a possibility of having poorly designed assessments which can lead to unsatisfactory competing rate of the intended learning outcomes of the course. For the previous obstacles, we propose a systematic approach to diminish such drawbacks. The proposed approach is used for generating automatically course exams, quizzes, exercises, and homework using Bloom's taxonomy. Furthermore, the proposed approach divides the total mark of the exam over the selected questions in the exam based on predefined criteria.

This paper is structured as follow. The next section presents a number of existing tools that are proposed to generate exams automatically. Then, the proposed approach to generate examination automatically is discussed. Furthermore, a list of requirements and the conceptual framework are also presented. Next, the implementation and the developed prototype are discussed. Finally, the paper is concluded and future directions are presented.

2 LITERATURE REVIEW

This section reviews related work dealing with generating exams out of question bank automatically.

There are different attempts conducted to consider the Bloom's taxonomy for generating exams automatically. For instance, the work presented in (Kale & Kiwelekar, 2013) considers four constraints to generate the exams. The constraints are proper coverage of units from course's syllabus, coverage of difficulty levels of the questions, coverage of cognitive levels of Bloom's taxonomy and the distribution of the marks across questions. Such constraints are considered for developed algorithm to generate the final paper exam. Another interesting work for classifying questions according to Bloom's taxonomy is presented in (Omar et al., 2012). The proposed work is a rule-based approach. However, the generation process of exams is not considered in this work.

Other approaches are related to the use of Natural Language Processing (NLP) to classify questions and assign weight for each question. For instance, authors in (Jayakodi et al., 2016) shows promising results in using NLP techniques to weight questions according to Bloom's taxonomy cognitive levels. Other researchers (Mohandas et al., 2015) propose the use of Fuzzy logic algorithm for the selection process of the questions depending on difficulty level.

Different tools were developed to validate the proposed approaches in the context of automatic exam generation. For instance, (Cen et al., 2010) presented a tool using J2EE tools to support educators by identifying the subject, questions types, and difficulty level. Accordingly, the proposed prototype will generate the exam in MS document format. The proposed work does not map questions to the course syllabus and Bloom's taxonomy. Other researchers (Gangar et al., 2017) proposes a tool which categorizes questions as knowledge-based, memorybased, logic-based, or application-based. The work uses a randomization algorithm for selecting questions from the question bank database. Furthermore, exams can be generated only for unit exams or final semester exams.

More comprehensive review of proposed approaches and tools for generating exams automatically are presented in (Joshi et al., 2016; Tofade et al., 2013; Taqi & Ali, 2016).

3 AUTOMATIC EXAM GENERATION APPROACH

Considering the different obstacles and challenges related to the assessments in a course module, the proposed approach provides a platform for selecting questions depending on ILOs and distributing marks based on specific criteria. The proposed approach for generating the exam is mainly based on Bloom's taxonomy. This enables the system to standardize the assessment of any course to a great extent. This is achieved by assigning the learning topic (contents), which can be a section of a chapter in a specific textbook, a video, or audio to corresponding ILO. Furthermore, also the questions related to each learning topic are assigned to the corresponding ILO.

In the proposed approach, the educator is responsible for defining a question and map it to a predefined ILO explicitly. The advantage of this approach is that it gives control to the educator. On the other hand, this can be a disadvantage in the way that it can take quite some time for the educator to do the mapping process between the learning topics, questions and the ILOs. However, supporting educators with an appropriate and usable tool can overcome this issue. Also, the manual approach can be complemented with classification algorithms to map topics and questions to related ILO automatically (Jayakodi et al., 2016).

The next sections presents the requirements for generating exams based on ILOs. Then, a conceptual framework (models and principles) is presented. Finally, the algorithm for generating the exams and distributing grades is explained.

3.1 Requirements

Based on the reviewed literature (Mohandas et al., 2015; Tofade et al., 2013; Alade & Omoruyi, 2014; Joshi et al., 2016; Omar et al., 2012), a number of requirements are derived to be considered in developing of an automatic exam generator. The requirements are as follow:

Question Variety: this requirement is mainly considered to provide different types of questions mapped to an ILO. This is achievable by providing different types of questions, both subjective and objective questions, e.g., essay questions, multiple choice, true/false, match column, multimedia questions, fill in blank, etc. that are related to a specific learning concept or topic.

Randomization: this requirement is used to guarantee that the generated exam does not have repeated or biased questions. It can be realized by means of random algorithms (Marzukhi & Sulaiman, 2014).

Educational Taxonomy Mapping: this requirement is considered to map a learning outcome to both a question and a learning topic. This will enable the educators to know the covered ILOs in each exam. Theretofore, the revised version of Bloom's taxonomy is considered in this research work.

Marks Distribution: there is a need to consider a fair distribution of the exam total mark over the composed questions. One way to achieve this is to use educators' experience to give score for each question manually. Other approach uses algorithms that consider the ratio of required time to solve the question (defined by the educator) and the specified time for the *exam* in general (defined by the educators). This is a simple approach for marks distribution for different questions in the exam.

ILO Validation: this requirement is mainly used for validating the defined ILOs according to Bloom's taxonomy. *This* is done by considering some keywords from a specific level (Remember, Understand, Apply, Analyze, Evaluate, Create). In other words, matching algorithms can be used to find the keyword from ILO and match it with a corresponding cognitive level from Bloom's taxonomy. For instance, a defined ILO can be "*explain the concept of object oriented programming"*. This ILO is related to the second level of the revised Bloom's taxonomy (Krathwohl, 2002), which is the understanding level. As a result, the validation algorithm starts searching for the action verbs inside the statement of the defined ILO ("explain" in the given example) and map it to the corresponding level of the Bloom' taxonomy.

Other requirements such as the security issues, usability aspects like ease to use, and ease to understand, are also considered in this work partly. However, there is still a need for evaluating the proposed prototype.

3.2 Conceptual Framework

In general, to be able to generate an exam by considering a number of parameters such as a number of selected topics, selected ILOs, exam time, etc., there is a need to maintain all required information in different models. In this approach, generating an exam depends on *Course Model, User Profile, ILO Model, Question bank, Generated Exam repository* and *the Generator Engine* (see Figure 1).

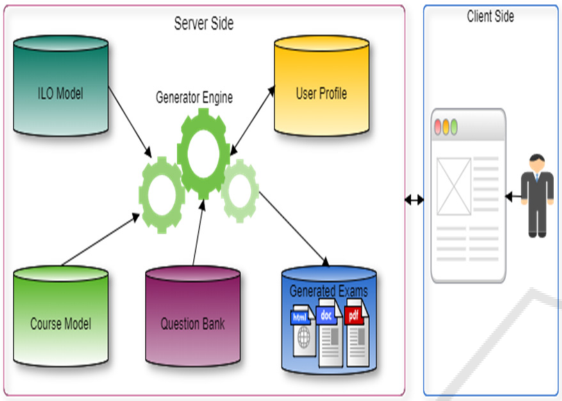

Figure 1: Conceptual Model for Generating Exams Automatically.

The *Course Model* is used to describe the different topics that will be covered in the course. Each topic is mapped to the related ILO (from the *ILO Model*). A topic can be related to only one ILO. However, an ILO can be related to different topics in the same course with a specific percentage.

The *User Profile* is used to maintain the educator's information such as his user name and password, taught courses, and created questions.

The *ILO Model,* as mentioned earlier, allows to associate questions to the different predefined ILOs of the course and this is an important step to assess learning process depending on familiar standards such as Bloom's taxonomy. Therefore, a repository of ILOs is required to hold the information about each ILO such as Bloom's taxonomy level, related course name, related learning topic, covered percentage of the ILO in the learning topic, and related questions.

The *Question Bank* is required to map each question to learning topic. It is important to mention that each ILO should be mapped to at least one question since ILO can be evaluated by different questions types. This mapping is important to help the educators in knowing the covered percentage of specific ILO in the exam. Furthermore, this will enable educators to keep track of covered ILOs in the course at a specific moment.

It is important to mention that the generated exam can be an electronic or paper-based exam. Electronic exams can be used for e-learning applications such as MOOCs where the questions can have multimedia contents such as animation, 3D models, simulation model, video, audio, etc. Therefore, the generated file is an XML format attached with different multimedia resources. On the other hand, paper exams can be generated in two formats MS-Word document or PDF files for use in classical courses.

The *Generated Exam Repository*: a repository of generated questions is used to store historical information of used questions in different exams, semesters, years, etc. Such information can be used by the educators to explore the previous generated exams.

In general, generated exams types, which are considered in the generation process, are quizzes, exercises, first exam, second exam, midterm exam, and final exam. Such assessments are used in different universities for different programs such as Engineering, Science, Business, Medical, etc.

The kernel of the framework is the *Generator Engine* which is responsible of realizing the creation of the different exams based on IF-ACTION rules. The IF-part of the hard coded rules contains three parameters: *learning topic*, *Bloom's taxonomy level*, and *required time* for solving the question. The ACTION-part of the rule uses a random algorithm for selecting a question out of the filtered questions based on the IF-part of the rule. Moreover, the generator engine is responsible for calculating the marks of each question in the exam depending on timing criteria. More details about the process of filtering, selecting, grading questions are presented in the next section.

3.3 Exam Generating Algorithm

To determine a question in an exam, the list of learning topics, which will be evaluated, need to be defined. This will narrow the possible questions that will be used for the generation process of the exam. The second level of categorization is related to the ILO that will be examined in the selected exam. This will narrow the sample of the possible questions from the previous step. Accordingly, the algorithm will start the selection of the question in a specific sequence from the selected topics till the final topics that are included in the exam. However, selecting a question related to specific topic and ILO is done as a random selection of the questions.

The mark for a selected question is dependent on the exam itself such as first, second, midterm, final, quiz, exercise, etc. For instance, a question can have 10 marks in a first exam which has relatively long time to be finished, but it can also have 5 marks in a quiz which has only a short time to be completed. Depending on a number of studies, there is a correlation between the time spent to complete the exam and the final grade that the student get at the end of the exam (Beaulieu & Frost, 1994; Landrum & Carlson, 2009; Kale & Kiwelekar, 2013). Similarly, our approach is considering the time specified by the educator to complete a specific question as an indicator for the score of the question. In other words, Question Mark = (ETQ / ETE) X (EM) where ETQ is the estimated time for the question, the ETE is the estimated time for the exam in total and EM is the exam grade in total.

$$Mark = \frac{EQT}{ETE} \times EM$$

Equation 1: Calculate the mark of the question.

Following the previous steps in the algorithm, our proposed algorithm satisfy the idea of generating a balanced and sequenced questions approach (Tofade et al., 2013; Susanti et al., 2015) as it sorts the selected questions depending on Bloom's taxonomy. Therefore, the questions that are mapped to a lower level of the cognitive level in Bloom's taxonomy such as remembering, understanding, and applying are placed in the first part of the exam. On the other hand, questions that are mapped to advanced level of the cognitive level from Bloom's taxonomy such as analyze, evaluate, and create can appear later in the exam. According to psychologists, this will create a safe environment as first the students are asked a couple of simple questions and then the students are involved in the more analytical questions.

4 IMPLEMENTATION

To validate our proposed solution for automatic generation of exams, we have developed a web-based prototype using PHP[1] and MySQL[2] running on Apache Tomcat[3]. To be able to handle the question bank, a server stores questions and related data such as ILOs and corresponding learning topics in the database.

As a first step, the educator needs to enter the details about the course so that he can enter the course name, topics to be covered in the course, ILOs and their corresponding cognitive level in Blooms' taxonomy. Validation of the ILO and the corresponding ILO is done at runtime. As a result a notification message will be displayed to the educator if there is misleading information.

After that, question entry is the next step. Each question is added manually using the developed prototype. As depicted in Figure 2, a question can be added to the database by specifying the course name, related ILO, corresponding dimension of Bloom's Taxonomy, an expected time for solving the question in minutes, and the question type. After specifying the question type, the educator will be shown a GUI to enter the question, the options, URL for multimedia contents (for electronic exams) and the correct answer. As mentioned earlier, there are a number of question types such as True/False, Essay, etc. Accordingly, the GUI will depend on the option that the educator select for the type of the exam.

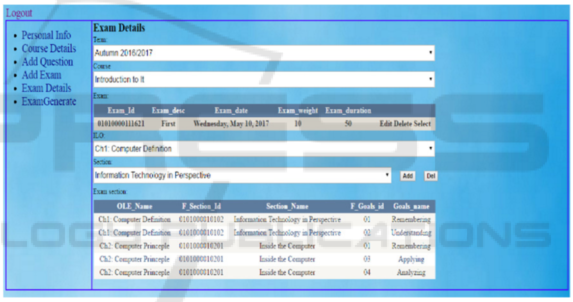

Figure 2: Exam Details Screen.

Other type of questions is related to providing images along with a textual question and the correct answer. Obviously, generated exams with multimedia contents such as animation, 3D models, etc. can be used only in electronic exams rather than paper-based exams.

After entering the required information about the question, the educator will create an exam by specifying the following data: name (such as First, Second, etc.), max grade, time to complete the exam, the semester and the exact date and time for conducting the exam. After filling in the required information about the exam, the educator will be directed to a new screen which asks him to enter the

[1] https://secure.php.net
[2] https://www.mysql.com
[3] http://tomcat.apache.org

content of the exam such as the sections and ILOs to be included in the exam. See Figure 3.

Finally, the educator can move to the generation screen which will display a list of selected questions based on the proposed algorithm. The educator needs to specify the type of the generated exam such as XML (for electronic exams), PDF, or Document.

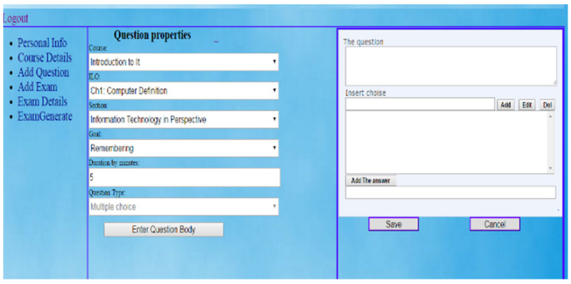

Figure 3: Adding a question to the database screen.

There are a number of limitations in the current proposed prototype. One of the limitations is that the educators are not able to modify or update any generated exam. This functionality can be important to allow the educator to change a specific question or select manually other alternative questions for a specific ILO. Another limitation is related to the few number of the questions stored so far in the database.

5 CONCLUSION AND FUTURE WORK

Traditional preparation of exams is considered as a tedious process, difficult to track all topics according to the syllabus, requiring a high mental demand to avoid question repetition and to avoid questions that are too easy or too tough. The proposed prototype addresses the above-mentioned obstacles in an effective way by generating exams automatically based on Bloom's taxonomy. The proposed work facilitates generating exams automatically depending on the intended learning outcomes of a course module.

As this paper presents the general goal of our research, there are a couple of research extensions to be considered in the future. To improve the alignment of assessment with learning outcomes, the next step is to classify different questions, that can be retrieved from a Learning Management System (LMS), automatically using some sort of classification algorithms such as Support Vector Machine (SVM), Naïve Bayes (NB), and k-Nearest Neighbour (k-NN) or combine these algorithms (Al-smadi et al., 2016; Abduljabbar & Omar, 2015).

Usability of any automatic exam generator system could be a problem as the graphical user interface and different considered aspects such as ILO, learning topic, exams time, etc. could become relatively complex. We are planning to conduct an experiment on improved version of the prototype to validate the issue of usability and acceptability of the system. The evaluation will be mainly based on ISONORM 9241/110-S Evaluation Questionnaire (ISONORM), Subjective Impression Questionnaire (SIQ), Qualitative Feedback (QF), and Workload Perception (WP) to validate different aspects such as easy to use, easy to understand, mental demand, etc.

Another important improvement, that will be conducted, is to support educators with a visualization tool for viewing easily the covered ILO in all generated exams for a specific course. This will be helpful in monitoring and tracking the covered ILOs so that missing ILOs can be included in the future exams of the course.

Considering standards for assessments such as IEEE Learning Object Metadata, IMS Question and Test Interoperability, etc. will be investigated in next stage of this research work to enhance the work from two points of view. First, it will facilitate the automatic mapping process between Bloom's Taxonomy and questions. Such standards can be used to import questions from existing question bank that are part of many Learning Management Systems (LMS) such as Moodle, Blackboard, Canvas, etc.

REFERENCES

Abduljabbar, D. A. & Omar, N. (2015). Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination. *Journal of Theoretical and Applied Information Technology*. 78 (3). pp. 447–455.

Al-smadi, M., Margit, H. & Christian, G. (2016). *An Enhanced Automated Test Item Creation Based on Learners Preferred Concept Space*. 7 (3). pp. 397–405.

Ala-Mutka, K. M. (2005). A Survey of Automated Assessment Approaches for Programming Assignments. *Computer Science Education*. 15(2), pp.83–102.

Alade, O. M. & Omoruyi, I. V. (2014). Table Of Specification And Its Relevance In Educational Development Assessment. *European Journal of Educational and Development Psychology*. 2 (1). pp. 1–17.

Beaulieu, R. P. & Frost, B. (1994). Another Look at the Time-Score Relationship. *Perceptual and Motor Skills*. 78 (1). pp. 40–42.

Biggs, J. B. & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Taxonomy of educational objectives: the classification of educational goals. B. S. Bloom (ed.). Longman Group.

Blumberg, P. (2009). Maximizing learning through course alignment and experience with different types of knowledge. *Innovative Higher Education*. 34 (2), pp 93–103.

Cen, G., Dong, Y., Gao, W., Yu, L., See, S., Wang, Q., Yang, Y. & Jiang, H. (2010). A implementation of an automatic examination paper generation system. *Mathematical and Computer Modelling*.51(11), pp. 1339-1342.

Fuller, U., Riedesel, C., Thompson, E., Johnson, C.G., Ahoniemi, T., Cukierman, D., Hernán-Losada, I., Jackova, J., Lahtinen, E., Lewis, T. L. & Thompson, D.M. (2007). Developing a computer science-specific learning taxonomy. *ACM SIGCSE Bulletin*. 39(4), pp. 152-170.

Gangar, F. K., Gori, H.G. & Dalvi, A. (2017). Automatic Question Paper Generator System. *International Journal of Computer Applications*. 66 (10). pp. 42–47.

Jayakodi, K., Bandara, M. & Perera, I. (2016). An automatic classifier for exam questions in Engineering: A process for Bloom's taxonomy. In: *Proceedings of 2015 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2015*. 2016.

Joshi, A., Kudnekar, P., Joshi, M. & Doiphode, S. (2016). A Survey on Question Paper Generation System. In: *IJCA Proceedings on National Conference on Role of Engineers in National Building NCRENB*. 2016, pp. 1–4.

Kale, V. M. & Kiwelekar, A.W. (2013). An algorithm for question paper template generation in question paper generation system. In: *2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering, TAEECE 2013*. 2013, Konya, p. 256–261.

Krathwohl, D.R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*. 41(4), pp.212-218.

Landrum, R.E. & Carlson, H. (2009). The Relationship Between Time to Complete a Test and Test Performance. *Psychology Learning & Teaching*.

Manuel Azevedo, J., Patrícia Oliveira, E. & Damas Beites, P. (2017). How Do Mathematics Teachers in Higher Education Look at E-assessment with Multiple-Choice Questions. In: *Proceedings of the 9th International Conference on Computer Supported Education*. 2017. pp. 137-145.

Marzukhi, S. & Sulaiman, M.F. (2014). Automatic generate examination questions system to enhance preparation of learning assessment. In: *2014 4th World Congress on Information and Communication Technologies, WICT 2014*. 2014. pp. 6-9.

Mohandas, M., Chavan, A., Manjarekar, R. & Karekar, D. (2015). Automated Question Paper Generator System. *International Journal of Advanced Research in Computer and Communication Engineering*. 4(12), 676-678.

Omar, N., Haris, S. S., Hassan, R., Arshad, H., Rahmat, M., Zainal, N.F.A. & Zulkifli, R. (2012). Automated Analysis of Exam Questions According to Bloom's Taxonomy. *Procedia - Social and Behavioral Sciences*.59 pp. 297 – 303.

Susanti, Y.., Iida, R.. & Tokunaga, T.. (2015). Automatic generation of english vocabulary tests. *CSEDU 2015 - 7th International Conference on Computer Supported Education, Proceedings*.pp.77-87.

Taqi, M. K. & Ali, R. (2016). Automatic question classification models for computer programming examination: A systematic literature review. *Journal of Theoretical and Applied Information Technology*. pp. 360–374.

Tofade, T., Elsner, J. & Haines, S. T. (2013). Best practice strategies for effective use of questions as a teaching tool. *American Journal of Pharmaceutical Education*. 77(7).pp. 55.