# Artificial Intelligence System in Aid of Pedagogical Engineering for Knowledge Assessment on MOOC Platforms: Open EdX and Moodle

https://doi.org/10.3991/ijet.v18i05.36589

Younes-Aziz Bachiri[1](✉), Hicham Mouncif[2]
[1]FST Beni Mellal, Sultan Moulay Slimane University, Beni Mellal, Morocco
[2]Multidisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco
younes-aziz.bachiri@usms.ma

**Abstract—**The aim of this research is to provide a novel educational model with the goal of reducing the expenses associated with manual question production and meeting the demand for a continual supply of new questions on MOOC platforms such as Moodle or Open EDX. We considered integrating machine-learning methods with natural language processing in order to increase the number and validity of assessing questions. To accomplish this, we developed a system that generates multilingual questions automatically. Various kinds of evaluation were conducted with two factors in mind: evaluating MOOC learners' competency and the similarity of the generated questions to those created by humans. The first evaluation is based on subjective judgment by three MOOC creators, while the second is based on replies from MOOC participants on machine-generated and human-created questions. Both evaluations revealed that the machine-generated questions performed on par with the human-created questions in terms of evaluating skills and similarity. Moreover, the results demonstrate that most of the produced questions (up to 82 percent) enhance e-assessment when the new suggested technology is used.

**Keywords—**MOOC, e-learning, automatic question generation, machine learning, multiple choice questions, natural language processing

## 1 Introduction

Massive Open Online Courses (MOOCs) are becoming increasingly essential in attaining the Sustainable Development Goal 4 [1]. MOOCs are free online courses that allow for limitless enrollment and unrestricted access over the web. MOOCs provide interactive learning resources using videos, quizzes, rapid feedback, and assignments. Additionally, they facilitate social engagement through forums, real-time chat, and social media. Several institutions have offered MOOCs to their students during the COVID-19 epidemic to provide uninterrupted study from home [2].

MOOC platforms are used by businesses as well as by educational institutions such as high schools, universities, and higher education institutes. They are suitable for any company that considers the development of its employees to be a priority. These learning management systems aim to stimulate, promote, support, and personalize human learning and are used in face-to-face or remote interaction situations [3].

Teachers provide educational resources, such as lessons and exercises. These exercises allow the teacher to assess the learners' level of knowledge acquisition on each topic. They also allow learners to self-assess and to check whether they have assimilated a notion that they are supposed to master.

A significant characteristic of conducting a course with many learners is the impossibility of delivering non-automated or peer-reviewed grades and comments. Human tutors cannot follow up with each student individually and cannot evaluate and mark work in MOOCs, but the architecture must allow large-scale feedback and engagement. To satisfy the needs of many learners, MOOCs use computer-graded assignments. However, computer-based grading is sometimes restricted, unsatisfactory, and inadequate, as it does not allow for partial marks or thorough explanations of responses.

Multiple Choice Questions (MCQ) are the most widely used as e-assessment tools. Indeed, these questionnaires are made up of items for which the learner is offered a choice among pre-established response options. The correction of such questionnaires is very simple because it is only a matter of verifying the correspondence between the choices selected by the learner and the responses to these items. MCQ correction can be automated, which facilitates the assessment of an unlimited number of learners and promotes self-assessment of knowledge by these learners [4].

MOOC researchers, at the forefront of a new era of digital education, need to find the perfect way to ensure that learners have retained the information. One of the most effective ways to achieve this is to offer exams composed of multiple-choice questions automatically generated thanks to machine learning. The quiz function found in the majority of open-source LMSs is crucial for gauging a student's mastery of the material presented in an online class [5].

These tests allow us to determine whether our teaching methods or the design of eLearning courses are effective; in other words, to find out if they offer the best eLearning experience possible.

Questions are generated from sentences in the multilingual learning text material and are classified as either gap-fill or factual. The question remains: what is the optimal technique for producing multilingual multiple-choice questions and exporting them to multiple MOOC platforms?

The additional value of this research is that it integrates learning theories, pedagogical approaches, and the acquisition of necessary artificial intelligence capabilities by offering a pedagogical model capable of automating assessment on MOOC platforms.

The rest of the paper is organized as follows: in the second section, we will discuss many related studies in the fields of question generating systems, as well as the possibility of automatic question system assessment. We will discuss our suggested question generating system, its strategy, and capabilities in the third section. The fourth section describes the settings and data of conducted experiments. In the fifth section, we evaluate our method and compare the generated questions to those generated by other state-of-the-art systems. Additionally, we will compare produced questions to human made questions to ascertain the contribution of question generation systems. In the last sixth segment, we will draw conclusions and provide some recommendations for further study in this field.

## 2 Related works

Since question generation requires minimal human effort, various approaches have been developed in this field. Automatic Question Generation (AQG) research began in the 1970's. Nowadays, AQG is gaining traction because of the growth of MOOCs and the widespread adoption of e-learning technologies during the corona virus period.

Liu et al. employed natural language processing techniques to algorithmically generate test items for both reading and listening cloze items and used a collocation-based process to identify distractors [6].

Aldabe et al. offered an Automatic Question Generator for Basque language test questions. The information source for this question generator is linguistically analyzed by actual corpora provided in XML mark-up language [7]. Pino et al. proposed a technique for improving the quality of automatically generated cloze and open cloze questions used for evaluation in the ill-defined domain of English as a Second Language vocabulary learning by the REAP tutoring system [8]. Agarwal & Mannem proposed a method for automatically generating gap-fill questions. The algorithm locates informative phrases in the document and produces gap-fill questions from them by first blanking keys from the sentences and then determining distractors for these keys [9]. Bhatia et al. presented an approach for phrase selection using current test items from the web. The phrases are chosen based on a pattern derived from previous queries. They also presented a new method for producing named entity distractors [10].

Narendra et al. described a method that, given an English article, creates a set of cloze questions. The algorithm is split into three sections: sentence selection, keyword selection, and distractor selection [11]. Kumar et al. provide a method that uses machine learning and natural language processing to create educationally acceptable gap-fill multiple choice questions from educational literature [12]. Majumder & Saha proposed a unique method for choosing informative phrases from an input corpus for the creation of multiple-choice questions. The system employs a series of pre-processing processes such as phrase simplification and co-reference resolution [13]. Shah et al. ranked collected keywords using the Inverse Document Frequency (IDF) metric and generated distractors using a Context-Based Similarity method based on Paradigmatic Relationship discovery approaches. The algorithm is trained using a dataset derived from Wikipedia [14]. Satria & Tokunaga gave an in-depth examination of the assessment of English pronoun reference questions generated automatically by machines [15].

Santhanavijayan et al. presented an algorithm for the automated creation of multiple-choice questions on any domain provided by the user. Additionally, this algorithm created analogous questions to assess pupils' linguistic skills [16].

Y. Bachiri & Mouncif developed a technique that automatically creates questions from video captions. Each course concludes with an evaluation question, which is often a multiple-choice assessment of the student's comprehension of the video's content. In terms of evaluating competence and similarity, questions generated by machines performed comparably to those generated by humans [17].

Furthermore, there are numerous reviews that discuss various approaches to question generation.

Le et al. conducted a study of the state of the art in terms of techniques for building educational applications that utilize question generating. They found that, while there are several ways for automatic question creation, only a few educational systems utilizing question generation have been created and used in practical classroom situations [18].

Divate & Salgaonkar examined many automated question generating systems to determine why automated question generation remains appealing to researchers. The emphasis is mostly on the work of analyzing and evaluating alternative approaches and methodologies [19].

Ch & Saha conducted a comprehensive study of methods for automatically generating multiple-choice questions. They described a general process for an automated multiple-choice question generating system. Six steps comprise the process [20].

Amidei et al. led a review of the assessment methods utilized in AQG. Their research, based on a sample of 37 articles, demonstrated that the growth of systems has not been followed by parallel development of the techniques used to evaluate them [21].

Kurdi et al. provided an overview of the AQG community and its activities, described current trends and advancements in the field, emphasized recent innovations, and recommended areas for improvement and future possibilities for AQG [22].

Das et al. presented an overview of approaches for automatic question generation and evaluation using textual and graphical learning resources. The purpose of this study is to gather the most modern methods for producing and analyzing questions autonomously [23].

## 3 Methods: automatic question generation and MOOC integration

This section describes our architecture for creating multiple-choice questions and details how to implement each phase.

Artificial Intelligence and machine learning algorithms allow our system to produce a large number of quality quizzes and assessments in seconds, which are then integrated into the MOOC platform.

The general procedure outlined by Ch & Saha [20] is broken into many phases. Although the number of phases and the general approach vary significantly amongst systems, most systems adhere to a typical workflow.

As seen in Figure 1, the system is divided into six phases: We will describe each aspect of the approaches utilized to create our assessment system in the sections that follow:

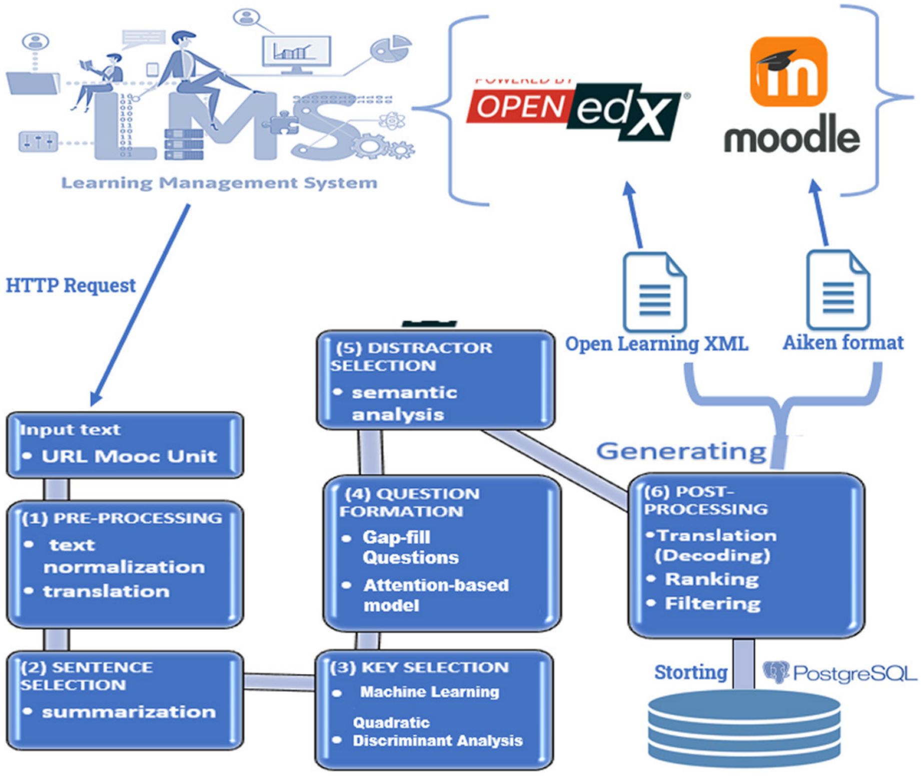


**Fig. 1.** Proposed architecture for automatic question generation

### 3.1 Pre-processing

Normalization of text refers to the process of converting incoming text to the appropriate format and removing superfluous information. Tokenization is the process of extracting tokens from text. There are a few notable exceptions to the requirement that a token be a single word. The process of lemmatization is the reduction of surface forms to their root forms.

Remove the suffixes. This is a more direct and expedient method than lemmatization. Sentence Segmentation: Separate text into sentences using the characters.,!, or? [24]

Text Language Identification is the process of guessing the language of a given text, whereas Text Translation is the act of converting a given text to another language. Frequently in multilingual MOOCs, we encounter situations where the text's language is unknown, or the language of the provided text document is altered to meet our demands.

We used Google Trans, an open-source Python library that provides access to the Google Translate API without any restrictions. Methods like detection and translation are called via the Google Translate Ajax API. [25].

### 3.2 Sentence selection

At times, input text may be overly long, preventing us from highlighting interesting passages. We chose to summarize to retain just the most pertinent sentences [26]. We used HOWSUMM, a novel large-scale dataset for the task of query-focused multi document summarization [27].

### 3.3 Key selection

It was necessary to train a binary classification model using one of the popular methods such as Logistic Regression, k-Nearest Neighbors, Decision Trees, Support Vector Machine, or Naive Bayes to determine if a word might be used as a response to our generated question [17].

**The training data set.** The model is trained using the Stanford Question Answering Dataset. Around 100,000 questions were produced from Wikipedia articles from the SQUAD v1 database [28].

We utilized spacy to describe the characteristics of the words. Once the text has been tokenized, we may access characteristics such as part of speech.

We utilized the Quadratic Discriminant Analysis classifier to determine whether a word is an answer.

**Quadratic Discriminant Analysis.** A quadratic decision boundary classifier is constructed by fitting class conditional densities to the input and applying Bayes' rule. Discriminant analysis is a term that refers to techniques that may be utilized for classification as well as dimensionality reduction. Linear discriminant analysis (LDA) is extremely popular since it functions as both a classifier and a tool for dimensionality reduction. Quadratic discriminant analysis (QDA) is a variation of linear discriminant analysis (LDA) that enables non-linear data separation. Finally, regularized discriminant analysis (RDA) is a hybrid technique that combines the advantages of LDA and QDA [29].

### 3.4 Question formation

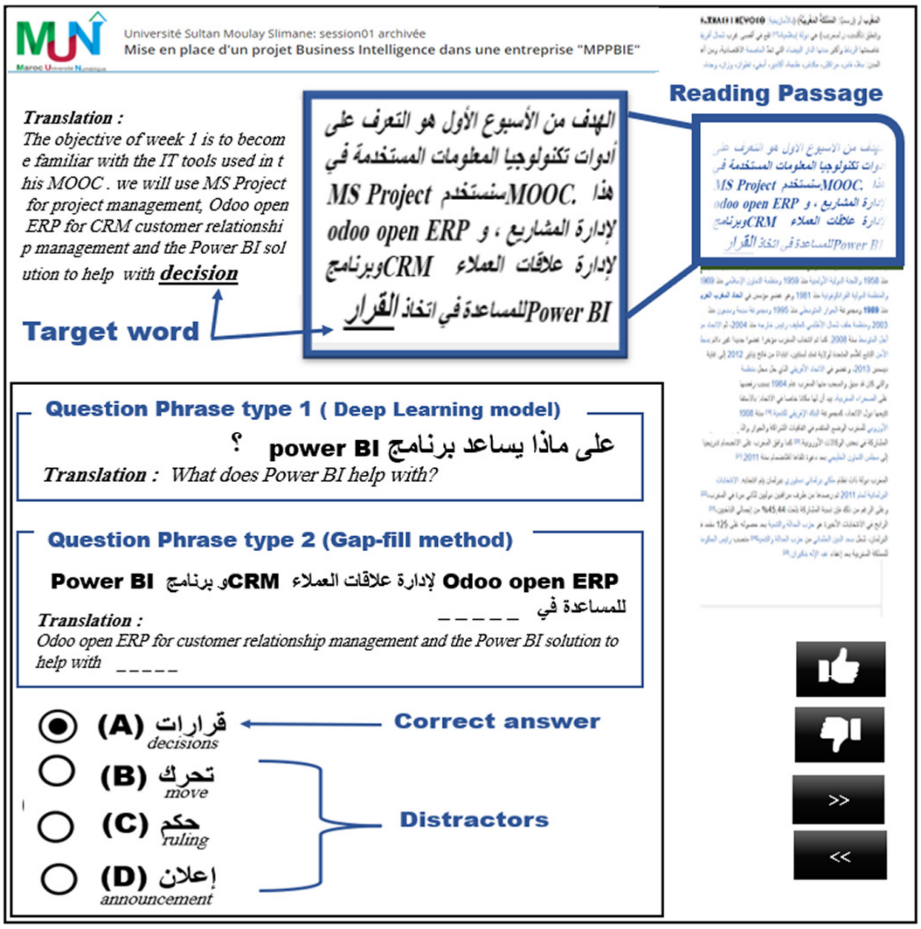
There are multiple alternatives for question formation, and we have adopted two techniques:

**Question phrase type 1 (based on transformers).** Using transformers, an attention-based model was adopted to generate questions automatically from given sentences and the target answer. The model can generate simple questions relating to unseen portions and responses averaging eight words in length.

We explored an automated question generating system that employs transformers instead of recurrent neural network (RNN). Our objective was to create questions using machine learning transformers that train faster and more effectively than RNNs. Instructors would benefit from saving time creating quizzes and examinations [30] [31].

**Question phrase type 2 (gap-fill questions).** We developed a method for automatically generating gap-fill questions. The algorithm analyzes the content for helpful phrases and creates gap-fill questions by first extracting keywords from sentences and then looking for distractors for these keywords [32].

Our AQG system in Figure 2 generates the following components: the stem (question), a target word which is the key answer, and a reading passage as a summary of the text and four distractors.



Université Sultan Moulay Slimane: session01 archivée
Mise en place d'un projet Business Intelligence dans une entreprise "MPPBIE"

*Translation :*
*The objective of week 1 is to becom e familiar with the IT tools used in t his MOOC . we will use MS Project for project management, Odoo open ERP for CRM customer relationshi p management and the Power BI sol ution to help with* ***decision***

الهدف من الأسبوع الأول هو التعرف على
أدوات تكنولوجيا المعلومات المستخدمة في
هذا MOOC.سنستخدم MS Project
لإدارة المشاريع ، و odoo open ERP
لإدارة علاقات العملاء CRMوبرنامج
Power BIللمساعدة في اتخاذ القرار

Question Phrase type 1 ( Deep Learning model)

على ماذا يساعد برنامج power BI ؟

*Translation : What does Power BI help with?*

Question Phrase type 2 (Gap-fill method)

Odoo open ERP لإدارة علاقات العملاء CRMو برنامج Power BI
للمساعدة في _ _ _ _ _

*Translation :*
*Odoo open ERP for customer relationship management and the Power BI solution to help with _ _ _ _ _*

● (A) قرارات *decisions*
○ (B) تحرك *move*
○ (C) حكم *ruling*
○ (D) إعلان *announcement*

**Fig. 2.** AQG sample: generated questions from reading passage

### 3.5 Distractor selection

Distractors of excellent quality are a critical component of the item and test development process. Multiple-choice questions should include as many options as feasible considering the item's content and the likelihood of distractors. Semantic analysis with the Word2vec tool was used [33] [8]. To control the level of question difficulty, distractors will then be replaced by the most common mistakes committed by students. The AI process will be enhanced by human intelligence.

### 3.6 Post-processing

Finally, post-processing improves the quality of the system generated MCQs, the system that produces MCQs may have various errors. These include incorrect punctuation, incorrect question terms, too long stems, numerical compatibility mistakes, and bad distractors. The system should reduce these mistakes.

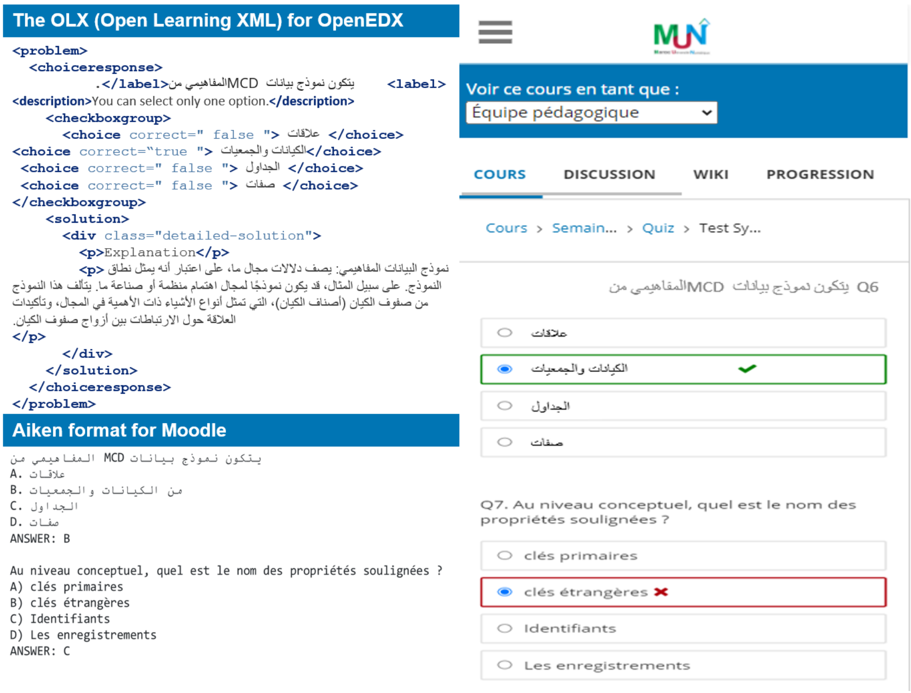
In this phase, the question, the result, and the distractors will be translated again into the original language and saved and stored into the database.

**Generating MOOC integration files.** This system phase attempts to process previously saved questions and generate the final question list. The system gives the examiner additional features, such as the ability to create a thorough test, generate complete questions, and finally export the quiz to multiple MOOC platforms such as Open EDX and Moodle.

*Open Learning XML.* The OLX (open learning XML) standard is the XML-based format for creating courses on the edX Platform. It can Transfer material between Open edX instances with OLX and create course content outside of edX Studio [34].

*Aiken format.* The Aiken format is a simple method for creating multiple choice questions in a text file in a human-readable format. (The GIFT format offers a greater number of options and may be less prone to errors, but it does not appear to be as straightforward as AIKEN.) The question must be fully contained on a single line. Each response must begin with a capital letter, followed by a period '.' or a bracket ')', and lastly by a space. Following that, the answer line must begin with "ANSWER:" and conclude with the matching letter [35].

Here is an illustration of how the two files are generated as shown in Figure 3:



The OLX (Open Learning XML) for OpenEDX

```
<problem>
  <choiceresponse>
          .</label>المفاهيمي منMCD يتكون نموذج بيانات      <label>
<description>You can select only one option.</description>
    <checkboxgroup>
      <choice correct=" false "> علاقات </choice>
<choice correct="true "> الكيانات والجمعيات</choice>
 <choice correct=" false "> الجداول </choice>
 <choice correct=" false "> صفات </choice>
</checkboxgroup>
    <solution>
      <div class="detailed-solution">
        <p>Explanation</p>
        <p> نموذج البيانات المفاهيمي: يصف دلالات مجال ما، على اعتبار أنه يمثل نطاق
النموذج. على سبيل المثال، قد يكون نموذجًا لمجال اهتمام منظمة أو صناعة ما. يتألف هذا النموذج
من صفوف الكيان (أصناف الكيان)، التي تمثل أنواع الأشياء ذات الأهمية في المجال، وتأكيدات
العلاقة حول الارتباطات بين أزواج صفوف الكيان.
</p>
      </div>
    </solution>
  </choiceresponse>
</problem>
```

Aiken format for Moodle

```
يتكون نموذج بيانات MCD المفاهيمي من
A. علاقات
B. من الكيانات والجمعيات
C. الجداول
D. صفات
ANSWER: B

Au niveau conceptuel, quel est le nom des propriétés soulignées ?
A) clés primaires
B) clés étrangères
C) Identifiants
D) Les enregistrements
ANSWER: C
```

Voir ce cours en tant que : Équipe pédagogique

Q6 يتكون نموذج بيانات MCDالمفاهيمي من

○ علاقات

● الكيانات والجمعيات ✓

○ الجداول

○ صفات

Q7. Au niveau conceptuel, quel est le nom des propriétés soulignées ?

○ clés primaires

● clés étrangères ✗

○ Identifiants

○ Les enregistrements

**Fig. 3.** Exporting the question list to Moodle and open EDX

# 4 Experiments

## 4.1 Experiment data

Two chapters from the MOOC "MPPBIE"[1] were used to measure acceptability. Each chapter is processed to yield two sorts of questions: those generated by the Transformer (type 1) and those generated by filling in the gaps (type 2). There were 303 questions created from these two units, 194 with the use of the attention-based model and 109 with the use of the gap-fill method.

## 4.2 Experiment settings

**Experiment 1: expert-based evaluation.** All produced questions were randomized and split into 3 test groups of around 101 questions each. To assess the performance

[1] The MOOC MPPBIE "IMPLEMENTATION OF A BUSINESS INTELLIGENCE PROJECT IN A COMPANY" is a course in management of information systems and corporate governance, published in the MUN platform Morocco Digital University. https://www.mun.ma/courses/course-v1:USMS+USMS001+session03/about

of our system, 3 human instructors are assigned to evaluate the acceptability of generated questions by determining whether each question phrase is acceptable and picking all alternatives that have the potential to persuade the evaluator to choose them as a response to the inquiry.

- Intolerable, the question is unsuitable for use in a real-world exam. Significant changes are required for real-world use.
- Tolerable, acceptable but can be improved: the question is acceptable as-is but might be enhanced further.
- Acceptable, the question can be used without modification in a real test.

**Experiment 2: assessing the learner's competency in a MOOC.** To evaluate the Django-based system used to power the Open EDX and Moodle platforms, 232 MOOC participants completed two quizzes.

In this experiment, we employed two distinct sorts of question sets:

- Test 1 has machine-generated questions (MQs) generated using the approach briefly explained in the "Automatic question generation" section by randomly picking 20 produced questions from the 140 already assessed as acceptable.
- Test 2 includes 20 human-made questions (HQs) taken from the official MOOC quiz.

## 5 Results and discussion

The most often used assessment technique is expert-based evaluation in which experts are given a sample of produced questions to examine. Given that expert assessment is a regular method for selecting questions for actual examinations, expert ranking is considered a reliable proxy for quality. However, it is critical to keep in mind that expert assessment gives just preliminary evidence regarding the quality of questions. Additionally, as we shall see later, the questions must be presented to a sample of students to ascertain their quality (empirical difficulty, discrimination, and reliability).

### 5.1 Expert-based evaluation

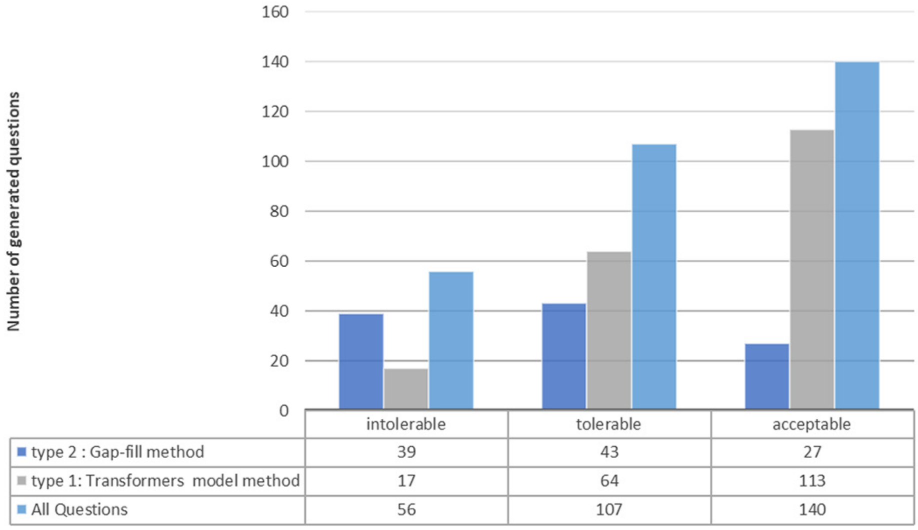
To begin with, invalid questions must be filtered out, and expert review is used to do this. Questions that are deemed invalid by experts (e.g., ambiguous, guessable, or not needing domain expertise) are filtered away. A well-chosen question set is critical for keeping MOOC participants engaged in question assessment motivated and interested in addressing these generated questions.

As shown in Table 1, 30% of the problematic generated questions were significantly impacted by the generic model's key selection. The used algorithm has a direct impact on teaching and learning results. However, 15% of the errors are grammatical; moreover, 10% of the errors are due to translation abnormalities, and 16% of the errors are due to fluency problems. Additionally, it was observed that some distractors had no relationship to the question or the actual response; this might be due to infrequent or specific locations in the Synset tree, as well as bad translation utilizing the Google Translate API.

**Table 1.** Frequency of evaluator observations

| Qualitative Observations | Evaluator 1 | Evaluator 2 | Evaluator 3 | Total |
|---|---|---|---|---|
| Correctness problem | 5 | 0 | 2 | 7 |
| Translation problem | 8 | 3 | 5 | 16 |
| Fluency problem | 2 | 16 | 8 | 26 |
| Semantic correctness problem | 3 | 2 | 3 | 8 |
| Discriminator quality problem | 9 | 10 | 14 | 33 |
| Learning outcome Key selection accuracy | 24 | 8 | 17 | 49 |
| Grammatical problem | 9 | 15 | 0 | 24 |

There are 56 question phrases that are considered intolerable by evaluators, which is 18% of all the 303 generated questions. From 2 generated question sets, the total acceptability rate is 82% and partial is 91% from questions formed by the Transformer Model. According to Figure 4, the acceptability rate for type 2 questions (gap-fill method) is low at only 64%.



| | intolerable | tolerable | acceptable |
|---|---|---|---|
| ■ type 2 : Gap-fill method | 39 | 43 | 27 |
| ■ type 1: Transformers model method | 17 | 64 | 113 |
| ■ All Questions | 56 | 107 | 140 |

**Fig. 4.** Expert-based classification of automatically generated questions

The vagueness of expert-based evaluation guidelines is another finding. For example, in an examination of reading comprehension questions, experts disagreed on whether reading the previous material is necessary to rank the question as excellent quality [36]. Researchers have also assessed question acceptability using scales with several categories (up to 9) but no clear categorization for each category. Zhang & Takuma discovered that reviewers use different scales and not all reviewers utilize all scales. We believe these two problems contribute to low expert inter-rater agreement [37]. To increase the accuracy of expert review data, researchers must clearly define the criteria used to assess issues. A pilot test with specialists is also required to validate the instructions and ensure that the instructions and questions are readily understood and comprehended by various responders.

### 5.2 Assessing the learner's competency in a MOOC

The primary objective of this evaluation is to determine if machine-generated questions are capable of accurately measuring MOOC learners' competency. We ask non-English-speaker MOOC participants (on the internet) to complete sets of machine-generated and human-created questions and then compare their results on the two sets to determine whether there is a correlation between them. We get in Table 2.

Analysis of the data is the act of gathering, summarizing, and analyzing data from test taker replies to determine the efficacy of individual question items. The difficulty index and discrimination index are two metrics that assist in determining the quality of multiple-choice questions used in an exam. Item analysis was conducted on 20 questions from both human-created questions (HQs) and machine-generated questions (MQs), and the results are summarized as indicated in Table 2.

**Table 2.** Metrics for comparing machine-generated versus human-created questions

| | Facility Index | Standard Deviation | Discrimination Index | Discriminative Efficiency |
|---|---|---|---|---|
| **Test1 (MQs)** | 17.00% − 77.53% | 37.75% − 50.25% | 5.60% − 52.20% | 7.37% − 64.98% |
| **Test2 (HQs)** | 15.31% − 81.63% | 36.19% − 50.26% | −13.37% − 55.73% | −16.65% − 67.62% |

**The facility index:** reflects the ratio of learners who properly answered the questions. The greater the facility index, the easier it is to answer the question.

**Discrimination index:** the relationship between a question's score and the total quiz score. That is, if you ask a good question, you expect that the students who correctly answer it will also correctly answer the other questions on the quiz. It is desirable to have a higher number.

**The discrimination efficiency:** it is a comparable report that attempts to avoid a problem. Questions with a discriminating percentage of ten or less have their question text highlighted in red. Discrimination efficiency works best for questions with facility indices ranging from 30 to 70. A perfect score is unlikely, although higher scores are preferable.

Equations for calculating the facility index (FI) and the discrimination index (DI) adapted from Moodle Statistics [38].

We also assess the efficacy of each question item by analyzing test taker replies using a statistical technique called item analysis. The item analysis employs two measures. One is the difficulty index, which indicates the percentage of test participants who properly answered the question item. Another indicator is the discrimination index, which reveals how effectively each question item may classify test participants according to their proficiency. Good question items have a reasonable difficulty index and a high discrimination index, indicating that they are neither too easy nor too tough, and are capable of discriminating test takers' ability.

For all these questions, the Facility index should be about the same. If some questions had substantially different Facility indices, this indicates that they were either significantly easier or significantly more difficult than the rest, implying that not all students received equally challenging examinations. The Discrimination index should ideally be high for all queries. Any question with a very low Discrimination index indicates that the question failed to differentiate student performance well. Figure 5 show that in both cases students have problems to deal with MPPBIE MOOC tests.

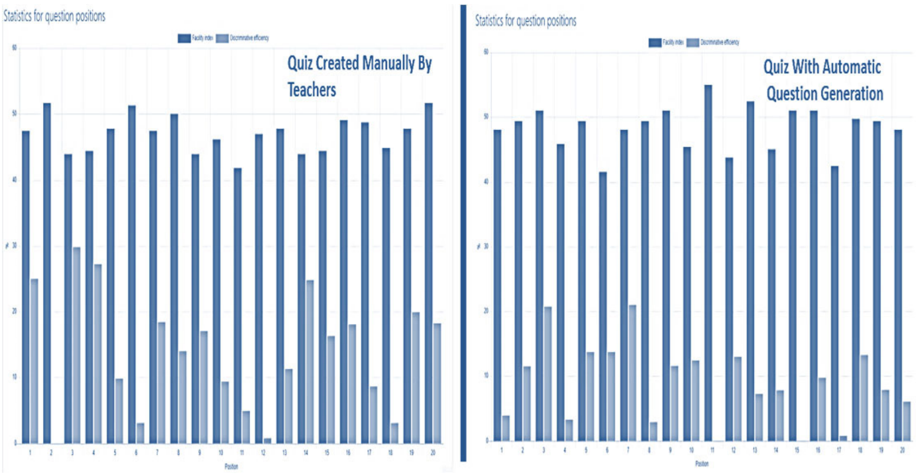


**Fig. 5.** Comparison between statistics of human made quiz vs AQG quiz

We assign the low percentage of question validation through testing with student cohorts to the time-consuming nature of these investigations and the ethical concerns associated with them. We must take care that these examinations have no impact on apprentices' grades or motives. For instance, MOOCs are built in such a way that anybody with an internet connection may enroll in courses for little or no cost. The characteristics of students who participate in assessments, such as their educational level and prior expertise with the subject being evaluated, are critical for study replication. Additionally, the characteristics of the individuals may explain the disparity in difficulty between studies.

Pandraju & Mahalingam provided a single model architecture employing "Text-to-Text Transfer Transformer" to generate questions from tables and text (T5) [39].

Liu et al. did a semi-automated evaluation of the nearby items' quality. 66.2 percent, 69.4 percent, 60.0 percent, and 61.5 percent of accurate phrases were created in response to the input request [6]. Aldabe et al. have an accuracy rating of more than 80% as determined by experienced language teachers [7]. Five English professors evaluated the length, simplicity, and difficulty level of sentences to get 66.3 percent accuracy [8]. Two biology students evaluated whether evaluation metrics was beneficial for learning and answerable. Evaluator-1 obtained 91.66 percent for phrase selection, 94.16 percent for key selection, and 60.05 percent for distractor selection, whereas Evaluator-2 obtained 79.16 percent for sentence selection, 84.16 percent for key selection, and 67.72 percent for distractor selection [9]. Bhatia et al. assessed by five domain experts obtained an average accuracy of 88 percent for distractors and 79.4 percent for keys [10]. Santhanavijayan et al. obtained an accuracy of 72 percent for informative phrases, 77.6 percent for blank generation, and 78.8 percent for distractor production [16]. Current advancements in pre-trained bidirectionally contextualized language models can also be included.

## 6 Conclusion

In this research, we presented a method for automatically creating multilingual questions. It can export a list of questions to the most popular MOOC platforms, such as Open EDX and MOODLE. It can also manage complicated texts and automatically locate appropriate distractors for MCQs.

The interesting aspect of the employed technique is its simplicity and modularity, which enables educators to spot areas of weakness and implement a solution.

To sum up, the first examination revealed a high correlation between the Machine Question test results and the Human Question test scores. Furthermore, outcomes may vary slightly when large groups of students, with varying learning cultures, for example, participate in the course and utilize the suggested technology. These variables should be evaluated in future studies.

Although the current study focuses on multiple-choice vocabulary questions, a potential future research area is to extend the system to create and evaluate other types of questions. Additionally, we explore varying the complexity of the vocabulary problems created automatically.

## 7 References

[1] P. Pradhan, L. Costa, D. Rybski, W. Lucht, and J. P. Kropp, 'A systematic study of Sustainable Development Goal (SDG) interactions', Earth's Future, vol. 5, no. 11, pp. 1169–1179, 2017. https://doi.org/10.1002/2017EF000632

[2] V. Kiketa et al., 'Design and implementation of a blended learning system for higher education in the democratic republic of Congo as a response to Covid-19 pandemic', International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 13, Art. no. 13, Jul. 2022. https://doi.org/10.3991/ijet.v17i13.30185

[3] V. Shurygin, N. Saenko, A. Zekiy, E. Klochko, and M. Kulapov, 'Learning management systems in academic and corporate distance education', International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 11, Art. no. 11, Jun. 2021. https://doi.org/10.3991/ijet.v16i11.20701

[4] D. Nicol, 'E-assessment by design: Using multiple-choice tests to good effect', Journal of Further and Higher Education, vol. 31, no. 1, pp. 53–64, 2007. https://doi.org/10.1080/03098770601167922

[5] R. Obeidallah and A. Shdaifat, 'An evaluation and examination of quiz tool within open-source learning management systems', International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 10, Art. no. 10, Jun. 2020. https://doi.org/10.3991/ijet.v15i10.11638

[6] C.-L. Liu, C.-H. Wang, Z. M. Gao, and S.-M. Huang, 'Applications of lexical information for algorithmically composing multiple-choice cloze items', in Proceedings of the second workshop on Building Educational Applications Using NLP, 2005, pp. 1–8. https://doi.org/10.3115/1609829.1609830

[7] I. Aldabe, M. L. De Lacalle, M. Maritxalar, E. Martinez, and L. Uria, 'Arikiturri: An Automatic question generator based on corpora and NLP techniques', in International Conference on Intelligent Tutoring Systems, 2006, pp. 584–594. https://doi.org/10.1007/11774303_58

[8] J. Pino, M. Heilman, and M. Eskenazi, 'A selection strategy to improve cloze question quality', in Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains. 9th International Conference on Intelligent Tutoring Systems, Montreal, Canada, 2008, pp. 22–32.

[9] M. Agarwal and P. Mannem, 'Automatic gap-fill question generation from text books', in Proceedings of the sixth workshop on innovative use of NLP for building educational applications, 2011, pp. 56–64.

[10] A. S. Bhatia, M. Kirti, and S. K. Saha, 'Automatic generation of multiple choice questions using wikipedia', in International conference on pattern recognition and machine intelligence, 2013, pp. 733–738. https://doi.org/10.1007/978-3-642-45062-4_104

[11] A. Narendra, M. Agarwal, and R. Shah, 'Automatic cloze-questions generation', in Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013, 2013, pp. 511–515.

[12] G. Kumar, R. E. Banchs, and L. F. D'Haro, 'Automatic fill-the-blank question generator for student self-assessment', in 2015 IEEE Frontiers in Education Conference (FIE), 2015, pp. 1–3. https://doi.org/10.1109/FIE.2015.7344291

[13] M. Majumder and S. K. Saha, 'A system for generating multiple choice questions: With a novel approach for sentence selection', in Proceedings of the 2nd workshop on natural language processing techniques for educational applications, 2015, pp. 64–72. https://doi.org/10.18653/v1/W15-4410

[14] R. Shah, D. Shah, and L. Kurup, 'Automatic question generation for intelligent tutoring systems', in 2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA), 2017, pp. 127–132. https://doi.org/10.1109/CSCITA.2017.8066538

[15] A. Y. Satria and T. Tokunaga, 'Evaluation of automatically generated pronoun reference questions', in Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, 2017, pp. 76–85. https://doi.org/10.18653/v1/W17-5008

[16] A. Santhanavijayan, S. R. Balasundaram, S. H. Narayanan, S. V. Kumar, and V. V. Prasad, 'Automatic generation of multiple choice questions for e-assessment', International Journal of Signal and Imaging Systems Engineering, vol. 10, no. 1–2, pp. 54–62, Jan. 2017. https://doi.org/10.1504/IJSISE.2017.084571

[17] Y. Bachiri and H. Mouncif, 'Increasing student engagement in lessons and assessing MOOC participants through Artificial Intelligence', in Business Intelligence, Cham, 2022, pp. 135–145. https://doi.org/10.1007/978-3-031-06458-6_11

[18] N.-T. Le, T. Kojiri, and N. Pinkwart, 'Automatic question generation for educational applications–the state of art', in Advanced computational methods for knowledge engineering, Springer, 2014, pp. 325–338. https://doi.org/10.1007/978-3-319-06569-4_24

[19] M. Divate and A. Salgaonkar, 'Automatic question generation approaches and evaluation techniques', Current Science, vol. 113, no. 9, pp. 1683–1691, 2017. https://doi.org/10.18520/cs/v113/i09/1683-1691

[20] D. R. Ch and S. K. Saha, 'Automatic multiple choice question generation from text: A survey', IEEE Transactions on Learning Technologies, vol. 13, no. 1, pp. 14–25, 2018. https://doi.org/10.1109/TLT.2018.2889100

[21] J. Amidei, P. Piwek, and A. Willis, ‘Evaluation methodologies in Automatic Question Generation 2013–2018’, presented at the Proceedings of The 11th International Natural Language Generation Conference, Tilburg, The Netherlands, Nov. 2018, pp. 307–317. Accessed: Feb. 07, 2022. [Online]. Available: http://oro.open.ac.uk/57517/; https://doi.org/10.18653/v1/W18-6537

[22] G. Kurdi, J. Leo, B. Parsia, U. Sattler, and S. Al-Emari, ‘A systematic review of automatic question generation for educational purposes’, Int J Artif Intell Educ, vol. 30, no. 1, pp. 121–204, Mar. 2020. https://doi.org/10.1007/s40593-019-00186-y

[23] B. Das, M. Majumder, S. Phadikar, and A. A. Sekh, ‘Automatic question generation and answer assessment: A survey’, Research and Practice in Technology Enhanced Learning, vol. 16, no. 1, p. 5, Mar. 2021. https://doi.org/10.1186/s41039-021-00151-1

[24] L. Bednarik and L. Kovacs, ‘Implementation and assessment of the automatic question generation module’, in 2012 IEEE 3rd international conference on cognitive infocommunications (CogInfoCom), 2012, pp. 687–690. https://doi.org/10.1109/CogInfoCom.2012.6421938

[25] M. O. Prates, P. H. Avelar, and L. Lamb, ‘Assessing gender bias in machine translation–a case study with Google translate’, arXiv preprint arXiv:1809.02208, 2018. https://doi.org/10.1007/s00521-019-04144-6

[26] A. Kurtasov, ‘A system for generating cloze test items from russian-language text’, in Proceedings of the Student Research Workshop associated with RANLP 2013, 2013, pp. 107–112.

[27] O. Boni, G. Feigenblat, G. Lev, M. Shmueli-Scheuer, B. Sznajder, and D. Konopnicki, ‘HowSumm: A Multi-Document Summarization Dataset Derived from WikiHow Articles’, arXiv preprint arXiv:2110.03179, 2021.

[28] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, ‘Squad: 100,000+ questions for machine comprehension of text’, arXiv preprint arXiv:1606.05250, 2016. https://doi.org/10.18653/v1/D16-1264

[29] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, ‘Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions’, Current Applied Physics, vol. 11, no. 3, pp. 740–745, May 2011. https://doi.org/10.1016/j.cap.2010.11.051

[30] K. Kriangchaivech and A. Wangperawong, ‘Question generation by transformers’, arXiv preprint arXiv:1909.05017, 2019.

[31] Y.-H. Chan and Y.-C. Fan, ‘A Recurrent BERT-based Model for Question Generation’, in Proceedings of the 2nd Workshop on Machine Reading for Question Answering, Hong Kong, China, Nov. 2019, pp. 154–162. https://doi.org/10.18653/v1/D19-5821

[32] Y. Bachiri and H. Mouncif, ‘Applicable strategy to choose and deploy a MOOC platform with multilingual AQG feature’, in 2020 21st International Arab Conference on Information Technology (ACIT), Nov. 2020, pp. 1–6. https://doi.org/10.1109/ACIT50332.2020.9300051

[33] I. Aldabe and M. Maritxalar, ‘Automatic Distractor Generation for Domain Specific Texts’, in Advances in Natural Language Processing, Berlin, Heidelberg, 2010. pp. 27–38. https://doi.org/10.1007/978-3-642-14770-8_5

[34] A. Chorana, A. Lakhdari, H. Cherroun, and S. Oulad-Naoui, ‘XML-based e-assessment system for Office skills in open learning environments’, Research and Practice in Technology Enhanced Learning, vol. 10, no. 1, p. 12, Jul. 2015. https://doi.org/10.1186/s41039-015-0008-y

[35] I. S. Mintii, S. V. Shokaliuk, T. A. Vakaliuk, M. M. Mintii, and V. N. Soloviev, ‘Import test questions into Moodle LMS’, arXiv:2010.15577 [cs], Oct. 2020, Accessed: Dec. 19, 2021. [Online]. Available: http://arxiv.org/abs/2010.15577; https://doi.org/10.31812/123456789/3271

[36] J. Mostow, Y.-T. Huang, H. Jang, A. Weinstein, J. Valeri, and D. Gates, 'Developing, evaluating, and refining an automatic generator of diagnostic multiple choice cloze questions to assess children's comprehension while reading', Natural Language Engineering, vol. 23, no. 2, pp. 245–294, Mar. 2017. https://doi.org/10.1017/S1351324916000024

[37] J. Zhang and J. Takuma, 'A Kanji learning system based on automatic question sentence generation', in 2015 International Conference on Asian Language Processing (IALP), Oct. 2015, pp. 144–147. https://doi.org/10.1109/IALP.2015.7451552

[38] S. H. P. W. Gamage, J. R. Ayres, M. B. Behrend, and E. J. Smith, 'Optimising Moodle quizzes for online assessments', International Journal of STEM Education, vol. 6, no. 1, p. 27, Aug. 2019. https://doi.org/10.1186/s40594-019-0181-4

[39] S. Pandraju and S. G. Mahalingam, 'Answer-Aware question generation from tabular and textual data using T5', International Journal of Emerging Technologies in Learning (iJET), vol. 16, no. 18, Art. no. 18, Sep. 2021. https://doi.org/10.3991/ijet.v16i18.25121

## 8 Authors

**Younes-Aziz Bachiri** is a doctoral candidate at the Laboratory of Innovation in Mathematics, Applications, and Information Technologies, Faculty of Sciences and Technology. Morocco's Sultan Moulay Slimane University, Beni Mellal. With over twelve years of experience instructing computer science in Moroccan secondary schools, he is currently pursuing a doctorate in education and artificial intelligence. He contributed to the realization of a number of MOOC programs and oversaw the university's open EdX platform. E-mail: younes-aziz.bachiri@usms.ma; ORCID: https://orcid.org/0000-0002-1834-9724.

**Hicham Mouncif**, Professor and Ph.D. Supervisor in the Department of Computer Sciences, Polydisciplinary Faculty of Beni Mellal, University Sultan Moulay Slimane, has already published many academic papers in distinguished journals based on teaching and research experience. Computer Systems Engineering Master's Coordinator His research interests include educational technologies, machine learning, transportation networking, and routing protocols. He is also a Director of Graduate Studies in Computer Science and Head of the Master of Informatics Systems Engineering (2019–present). E-mail: h.mouncif@usms.ma; ORCID: https://orcid.org/0000-0003-3312-8230.

Article submitted 2022-11-03. Resubmitted 2023-01-15. Final acceptance 2023-01-16. Final version published as submitted by the authors.