IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

Electronic Evaluation of Quality of Exams' Questions Written in Arabic Language Based on Bloom's Taxonomy

Adlan Balola Ali[1], Al shefa Abd algadir Hassan[2], Omar Balola Ali[3]

Sudan University for Sciences and Technology[1,2]

Bright Star University – El-Brega, Libya[3]

Abstract: This paper aims to analyse exam questions over four academic years in three colleges, to know levels of learning in final examination questions according to Bloom's levels of knowledge. The paper method used is quantitative and qualitative approach and content analysis. To collect the data, this study used documentation , 2318 questions were examined, taken from 110 final exams papers and from different levels, and the exam of the same course must be in all four years or three years at least , from 2015-2018. The analysis was carried out according to Bloom's classification of cognitive levels (knowledge - comprehension - application - analysis - synthesis - evaluation). There are two evaluation performed on this data. The first evaluation done by the experts and the second is electronic evaluation. The questions were distributed on all levels, but in different proportions, and the results were the following: The level of knowledge ranked first with 62.5%, the level of application in the second rank with 18.1%, and the level of comprehension in the third rank by 12.6%, while the percentage of the level of analysis was 3.9%, which is the fourth rank, and the level of evaluation was in the fifth rank by 1.6%, and the percentage of the level of synthesis was in the last rank. By 1.3%. There are no significant differences in the sample of the study in the level of knowledge between expert assessment and electronic assessment which is 0.49% , as well as there is no statistically significant differences in the level of comprehension which is 0.40%, and there is no statistical significance at the application level which is 0.84%, and there is no statistical significance at the level of analysis which 4.17%, and there are no statistical differences at the level of synthesis which is 2.22%, and there are no statistically significant differences in the evaluation level which is 3.85%.

Keywords: Bloom's Taxonomy, Exam, Natural Language Processing.

I. INTRODUCTION

Assessment plays an important part in the teaching-learning process at all levels of education. The main purpose of classroom assessment is to improve learning [1]. Since assessment plays such an important and significant part role in the future of students, there is no doubt that any assessment system will determine what and how students learn. Hence assessment will also determine what and how we teach. Education and assessment is related to each other. Without assessment, we have no other technique to know the significance or the effectiveness of learning experience to achieve the desired goal. Examination is one of the common methods to assess students' knowledge. Based-on the examination result, student's thoughts skills and behavioral can be developed [2].

The Arabic language [3] is one of the most common languages with more than 420 million speakers over the world. Unlike English, Arabic doesn't have upper cases. It also differs from other natural languages due to the presence of diacritics which represent a small vowel letters such as "fatha, kasra, damma, sukun, shadda, and tanween". The Arabic language's orthographic system is based on diacritics effect, where each specific type of diacritics produces different words with different meanings. This language has specific letters known as Arabic vowels (waw, yaa, alf) that require a special system of morphology and grammars. What also distinguishes Arabic is the huge amount of vocabularies and concepts.

Most of the text categorization systems have been developed for English language and just few of the developed systems were for Arabic language [4] . Because Arabic has a rich morphology, a complex syntax, complex semantics and very complex grammatical rules which distinguish it from other languages and make its learning, analysis and automatic processing difficult. The letters in the Arabic language are written in different forms based on their positions in the word. The letters may come in the front, middle, or last part of the word. To the best of my knowledge, the studies on Arabic language are very limited, in which there is a lack of Arabic corpus, language tools, and comprehensive studies on

DOI: 10.17148/IJARCCE.2023.12806

preprocessing Arabic texts. All these problems refer to diverse areas of challenges to categorize the specific Arabic textual data into a closed category.

This paper evaluate the exams questions using automatic classification of Arabic language exams questions by using identification from lexical and syntactical feature for each question. Many previous researches are only examined for English Language questions, so as according to one knowledge, there is no similar research in evaluating automatically Arabic Language questions. Thus, the researchers will make the dataset to answer the research questions. Feature extraction result of the questions will be classified by using rule-based approach applies Natural Language Processing (NLP).

In now a days, question classification is growing in popularity as it has an important role in question answering systems, information retrieval and it can be used in a wide range of other domains. The main aim of question classification is to accurately assign labels to questions based on expected answer type.

Based on Bloom Taxonomy, each exams' question must have at least one verb (keyword), where this keyword will determine the level of the question. In accordance, Bloom classified the keywords into six difficulty levels as described above. Using these concepts, this study will manipulate Bloom Taxonomy as principle to design a good examination question paper.

The study, based on the study questions and objectives, a rule-based approach applies Natural Language Processing (NLP) techniques is used in automatic classification of exams' questions using Bloom's cognitive level. The test items are a collection of examination questions from deferent subjects obtained from Computer Science, Sudan University and Omdurman University. The training model contains some semester examinations' questions from 2 years exams. Only written final examination question are taken for test items. The system will classify each of questions automatically to their corresponding verbs from the Taxonomy with the assistance of the developed rules. In order to determine the category of questions, this work excluded difficulty level of each question as a measuring factor.
The study uses descriptive analytical method which is based on the data collection and analysis of quality processes and information.

II. RELATED WORKS

The studies in the field of Blooms taxonomy in Arabic Language is very rare. However, there is a number of rich different studies conducted in English language using Bloom taxonomy classification of questions. The written exams questions are the most important tool in use among the evaluation tools that teachers use in universities or schools, and that importance increases when the results of those exams indicate that there is a shortage or lack of production, or that there are shortcomings or weaknesses in the production.

There is a number of studies have applied Bloom's taxonomy to automatic text classification in English language and there is a few effort in Arabic document classification. Some of the materials being classified, features and classifiers used in these studies are summarized below.

(Khorsheed et al. 2008) they presented a research in the classification in the Arabic language in order to measure the classification algorithms for documents written in the Arabic language.

They used a large database and various applications, such as email, web pages, and messages, automatic indexing of articles, searching for relevant information on the web and more. The results of different feature choices, weighting methods, and classification algorithms showed, on average, the superiority of the support vector machine, followed by the decision tree algorithm (C4.5) and Naı¨ve Bayes. The best classification accuracy was 97% for the Islamic subject's dataset, and the lowest accuracy was 61% for the Arabic poems dataset. [5]

(Othman et. al. 2018,) they used the HRWiTD document classification algorithm to classify Arabic texts of any type. It is a type of algorithm in artificial intelligence, and this algorithm divides the data into two parts, one part used for training the program on texts and another part for testing, and then the training part is used to classify the other file. The accuracy of the HRWiTD algorithm was evaluated after conducting a number of tests, and as a result, the accuracy of the HRWiTD algorithm was 86.84%. they used the same data for other algorithms, namely C5.0, KNN, SVM, NB, and C4.5, and the classification accuracy results were 52.86%, 52.38%, 51.90%, 51.90%, and 30%, respectively [6].

DOI: 10.17148/IJARCCE.2023.12806

(Qiao C. et. Al., 2018), their study is the first attempt to automatically analyze exams in students' Chinese analytical writing. Inspired by previous studies of text classification based on Bloom's taxonomy, this study attempts to classify the levels of cognition reflected in analytical writings. In particular, a rich set of lexical, syntactic, and semantic-level features are exploited when building classification models. Apply a feature selection process to filter out redundant features, which will further improve the performance of classification models [7]..

(Joe Harrison, et, el, 2017,) This research was carried out in the Netherlands in order to design exams automatically and also evaluate exams instead of the traditional method of developing exams and also methods of evaluating them. Developing traditional test questions and evaluating questions and comparing them to Bloom's Taxonomy can be time consuming if the test questions contain many questions of different types, and it can be easy to lose control of whether the questions in the assessment represent what is taught in the course material. In this paper, a machine learning technique is proposed to classify course questions and provides a clear overview of the chapters in Revised Bloom's Taxonomy found in these courses [8].

(Selvia et al, 2015) This study was conducted in the State of Indonesia in exams that were written in the Indonesian language. The study examines the extent to which the exam questions agree with the updated Bloom's classification. The selection is made on the basis of lexical features extraction and grammatical features extraction. The feature extraction output is classified using the Support Vector Machine (SVM) algorithm. The data used in the study are final exams in primary schools of the Indonesian Ministry of Education. This research showed that the proposed method can be used to classify Indonesian language question items well. The word in the Indonesian language is not always directly recognized, because sometimes the word has an affix. In order to know the label, there is a paste-delete process. An apostrophe is the added appendix that is attached to a word and gives new meaning [9].

(ABDULJABBAR et, al. 2015), This research was conducted in Malaysia and it is an attempt to use a new method to automatically classify exam questions according to cognitive levels of Bloom's taxonomy by implementing a combined strategy based on a voting algorithm that combines three machine learning classifiers. In this paper, a number of algorithms were studied and combined with each other, and these three classifications are (SVM, NB, and k-NN). The classification model achieved the highest score through the strategy of consolidation by applying mutual information, which proved to be promising and comparable with other similar models. These experiments are all in the right way to incorporate a number of artificial intelligence algorithms to classify exam questions [10].

(Choudhary T. et. al. 2014) this research investigated the impact of bloom’s taxonomy in introductory computer programming course to improve student’s learning experience and performance. They used text extraction and Text classification and Pearson’s Co-relation analysis performed using IBM SPSS tools to find out the relationship, if any, among the various levels of Blooms Taxonomy. As a result from the analysis the students is good in remembering level only, understanding level is not that much good, apply is. And in creating level h also the score is very less marks. So, overall needs to improve applying, understanding and creating level [11].

III. METHODOLOGY

A tool is implemented for Arabic text classification (ATC tool) in order to accomplish feature extraction and selection and can automatically split the dataset into training and testing sets. The size of these two partitions is determined by the user and also other tool created for this study to classify the verbs according to Bloom’s verbs.

In this work, a computer programming and some tools is adopted in classifying the question items into their corresponding Bloom’s cognitive level. The test items are a collection of examination questions in Programming subjects obtained from the Faculty of Computer Science, University of Sudan. The training set consists of 70 examination questions and the test dataset comprises of 30 questions as [12]. Only written final examination question are taken for test items. All of questions were manually categorized by a five expert in programming domain. The system will classify each of questions automatically to their corresponding verbs from the Taxonomy with the assistance of the developed rules.

In order to determine the level of questions, this work excluded the level of difficulty for each question as a measuring factor, and some verbs that fall into more than one level were also excluded and placed in the most used level for this verb.

The paper methodology depends on six main stages which are data collecting, questions extraction, preprocessing, feature extraction, classification and post-processing as shown in figure below.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

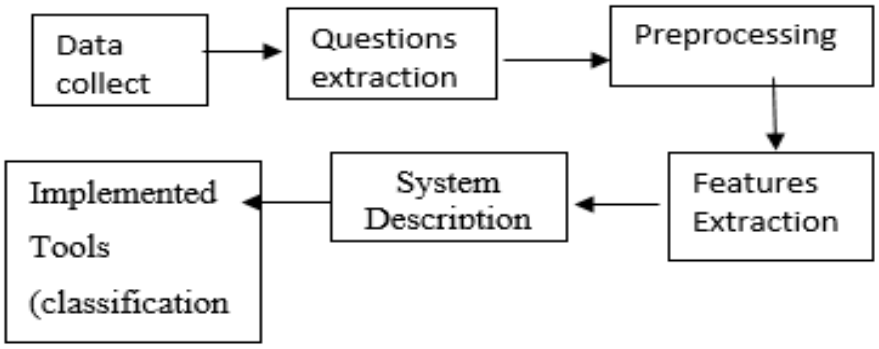


Fig. 1 Illustrates classification system

A. Preprocessing

In Arabic written text, some words are very common but add no additional meaning to the content of the text (such as في "in", على "on", أين "where"); these are called stop words [13].

The process of pre-processing is actually a process of improving the classification of text documents by removing the data that is worthless. The data may include worthless numbers, punctuations, kashida, Hamza ",," diacritics, and stop words. Some words do not belong to any classification such as prepositions, pronouns, etc., so we append them to a stop word list see Table 1. Preprocessing also normalize text documents by changing TaaMarboutah " ة" to " ا".

For example, they remove punctuation, diacritics, and non-Arabic letters, and specified Arabic letters are normalised: for instance, إ, أ, and آ are converted to ا (bare Alif); ي and, ئ (Yaa') are replaced by ي (Yaa'); and ة, ه (Taa', Haa') are replaced by ه (Haa'). Taghva [160] argued that removing the Hamza in this case does not affect the root. However, because of similarity in appearance, a more challenging type of spelling variation is that certain symbols joining Hamza or Madda with Alif (e.g., إ أ and آ) are sometimes written as a plain Alif (ا). Therefore ATC Tool is used to remove worthless data from the selective corpus.

TABLE 1 STOP WORDS

No.	Sentence	Arabic Words
1	Demonstrative pronouns	أولائك, هؤلاء, هتين, هتان, هذين, هذان, تلك, ذلك, هذه, هذا
2	Relative pronouns	الأسماء الموصولة
3	Numbers , Names	واحد ، اثنين ثلاثة .....
4	Prepositions of time and place	حروف الجر مثل من ، على وعيرها ظرف الزمان والمكان وغيرها من ادوات
5	Conjunction	التاء الواو, منذ, مذ , ربُّ, حتى, الكاف, اللام, إلى, الباء, في, على, عن, من,
6	Equations	الجمع والطرح والضرب وعمليات الجذور وغيرها من العلميات
7	Prepositions	لكن, لا, بل, أم, أو, حتى, ثم, الفاء, الواو

B. Feature Extraction

At this stage, the final exams questions are divided into two groups, a group of questions that are classified and processed by experts, and another group of final exam questions that are processed electronically based on the results of the test questions. The train group and the test group can be from an internal or external source. Features and word frequency list are extracted using the ATC tool.

The ATC tool lists and saves occurrences of each word in all specified train scripts in the train list file. It also lists and saves the frequency of each word in all test set texts in the test list file. In addition, add a field to train the list file and a

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

test list file to name the category of each word. The class of words in the train list is the actual class. On the other hand, the categories of words in the test list are set from the dataset learning file which contains the same words

Arabic Text Classification tool (ATC tool) was developed in Java to handle and process the dataset. The user interface for the ATC tool is shown in Fig. 2. The ATC tool incorporates the following main functions:

(a) Text preprocessing: This allows the user to remove numbers, punctuations, kashida and stop words and to normalize the texts by removing diacritics.
(b) Data division: This divides the dataset into two sets - one for training and the other set for testing. The user can manually specify text files to be included in either sets. Alternatively, the software can randomly assign those text files to either training or testing sets based on user selection of how much percentage of the whole dataset each set (training/testing) is.
(c) Feature extraction: This extracts and generates the frequency list of the dataset features (single words). The function can list and save the features frequency for the whole dataset, for a specific class or file, or for training/testing sets; taking into consideration user selection mentioned earlier.. In addition, the user can explore the frequency profile for certain list of words. The document frequency, relative frequency and relative document frequency of features can also be explored and saved.
(d) Feature selection: This calculates the importance of each feature locally (for each class) and globally (for all classes) based on 10 feature-selection methods.
(e) Data representation This generates the training and testing matrix elements where each element represents one selected feature from previous step.

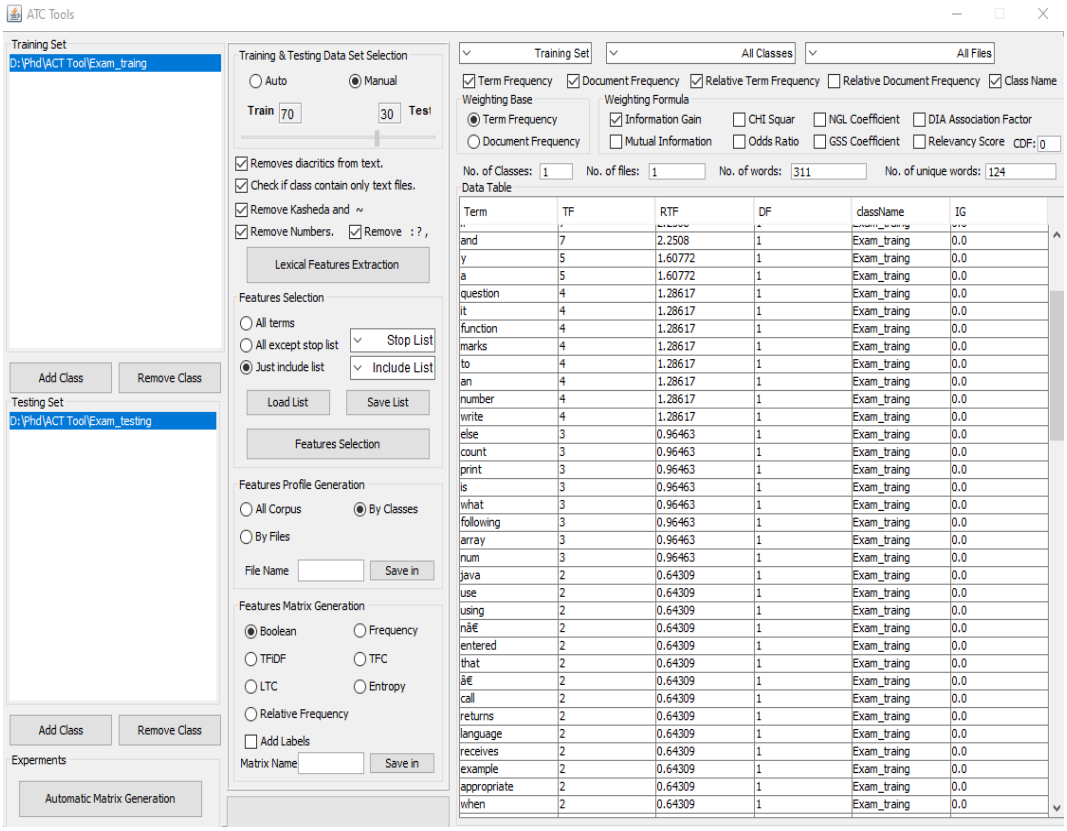


Fig 2 ACT Tool User Interface

C. System Description

As we mentioned previously, we should not depend on the verbs mentioned in the examination papers questions to know the classification entirely, but we can rely on it to somehow if the course teacher knows how to set the final exam questions and he is familiar with Bloom's classification and the levels of Bloom to put the right questions' verbs or write scenarios and what actions and how examination papers questions can be formulated.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

Depending on the specific scenario, the question may fall into more than one category in case of unfamiliarity with the type of questions and therefore, to overcome this problem, weights are assigned to the conflicting categories in the computer programs used and the weight is calculated based on the question category from the specialized experts. For example, based on the experts, it is possible to put weights in percentages for each verb and in which category this matter falls. For example, a specific command can be placed as 30% knowledge and 70% application, and based on these weights, the proposed system will choose the highest weight, i.e. 0.7 where the question will be categorized on It's an application, for example. Currently, we are still working on the training set to get optimized weights and patterns for each question. Fig. 3 illustrates the overall process for selecting a Bloom's Taxonomy category for a given question.

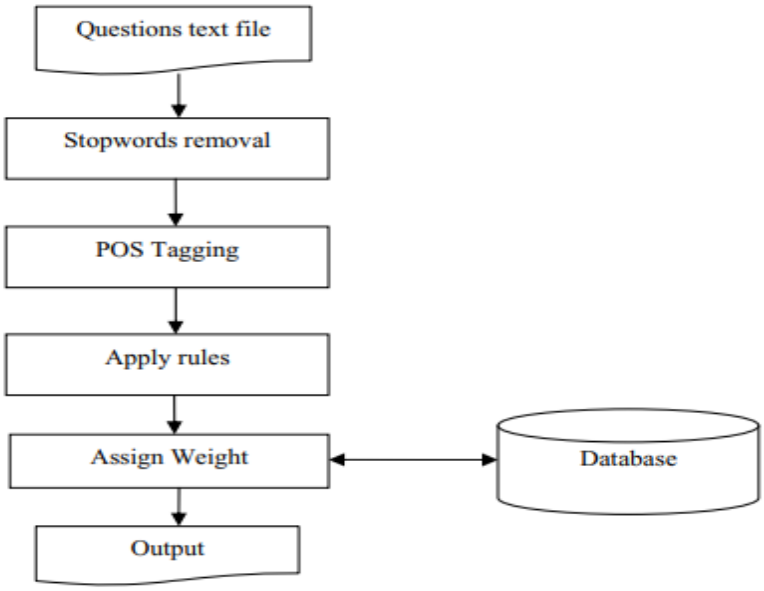


Fig. 3 Bloom's Taxonomy category

D. **Example of miss leading verbs**:

1) أكتب تعريفاً موجزا للحاسب الآلي
2) أكتب تعريفا موجزاً للذكاء الاصطناعي
3) أكتب تعريفا موجزاً للبريد الالكتروني

Lecturers who are not familiar to write a good quality exam questions may fall in a mistake by writing a misplaced verbs, therefore the question's category may changes from one classification to another, like the previous examples. The question begins with the verb "write" but the question asking to write a definition ,so the category of the question has been hanged from high level the application level to low level the knowledge, even though the question began with the verb "write" which indicates the application level. The right verb of these question should have been :define" and not "write" and so on.

أذكر الفرق بين RAM,ROM
أذكر الفرق بين البرنامج الهدف(object program) والبرنامج المصدر(source program)

The same as the previous problem also repeated in this question, but in another verb "mention" , this verb is in the level of remembering in Bloom's classification which is low level, although the question is about the difference between two things, and the verb "difference" in in the level of comprehension, so the level of the question changed from the level of comprehension to the level of knowledge and this problem was observed in many exam questions.
In the above question, there are more than one verb in the question and from different category levels, which reduces the possibility of identifying the type of verb level, and such a question is frequently asked in the questions of the chosen exams.

E. Implemented Tools

To experiment our method, we used the ATC data classification system and also we built a new system to identify the actions verbs according to Bloom's classification. The system was built using the Java programming language and some

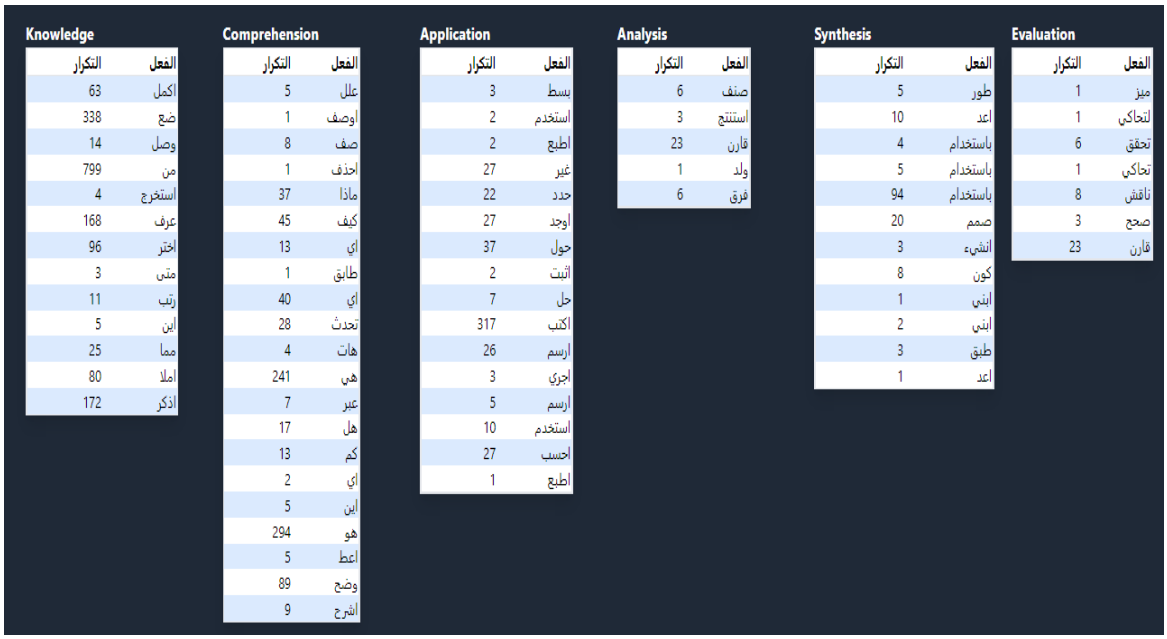
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

other software packages and tools. We tested the system using real data collected from final exam questions from Sudan University and Omdurman Islamic University.



Knowledge

الفعل	التكرار
اكمل	63
ضع	338
وصل	14
من	799
استخرج	4
عرف	168
اختر	96
متى	3
رتب	11
اين	5
مما	25
املا	80
اذكر	172

Comprehension

الفعل	التكرار
علل	5
اوصف	1
صف	8
احذف	1
ماذا	37
كيف	45
اي	13
طابق	1
اي	40
تحدث	28
هات	4
هي	241
عبر	7
هل	17
كم	13
اي	2
اين	5
هو	294
اعط	5
وضح	89
اشرح	9

Application

الفعل	التكرار
بسط	3
استخدم	2
اطبع	2
غير	27
حدد	22
اوجد	27
حول	37
اثبت	2
حل	7
اكتب	317
ارسم	26
اجري	3
أرسم	5
استخدم	10
احسب	27
اطبع	1

Analysis

الفعل	التكرار
صنف	6
استنتج	3
قارن	23
ولد	1
فرق	6

Synthesis

الفعل	التكرار
طور	5
اعد	10
باستخدام	4
باستخدام	5
باستخدام	94
صمم	20
انشيء	3
كون	8
ابني	1
ابني	2
طبق	3
اعد	1

Evaluation

الفعل	التكرار
ميز	1
لتحاكي	1
تحقق	6
تحاكي	1
ناقش	8
صحح	3
قارن	23

Fig. 4 Result of Classifying The Exam Questions

F. Experiments and Results of electronic evaluation

After conducting a comprehensive study and comparison to ascertain the validity of the automated assessment in order to compare it with the expert assessment on the final examination questions using the programs and tools mentioned above, some ideas and conclusions can be drawn on the validity or incorrectness of the assessment or the problems resulting from the automated assessment.

As expected, the difficulty of automatic identification of Arabic verbs and their classification according to Bloom's classification, due to the frequent use of the verb in different ways, the large number of derivations, and the failure of the exam authors to follow the scientific methods of the questions, which constitutes a great difficulty in automatic identification of the classification of questions.

Some action verbs in more than one cognitive level may mean that the verb of the question could be applied to more than one level, which results in the lack of recognition of the verb. In the following, we list the most important results.

statistical analysis of the verbs used in the examinations

A detailed statistical analysis for the structure of questions is presented in Tables 2, and Table 3 which shows the percentage of expert evaluation for the final exams questions and electronic evaluation for the final exams questions throughout the duration of the research (from the year 2015 and 2018).

This level of analysis indicated that the deference between the experts evaluation of the questions and the electronic evaluation for the questions across the duration of the project show small deference.

statistical analysis of the cognitive levels

The following table demonstrates the frequencies and percentage of the distribution of cognitive levels of the Bloom's Taxonomy in the final exams questions and showed the electronic evaluation and the expert's evaluation and the deference between them.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

TABLE 2 COMPARISON DISTRIBUTION OF VERBS IN THE FINAL EXAMS QUESTIONS FOR ALL THE THREE COMPUTER DEPARTMENTS.

No	Arabic Verb	cognitive Level	Expert Evaluation	Electronic Evaluation	The deference
1	قارن	Analysis	24	23	1.00
2	استنتج	Analysis	3	3	0.00
3	فرق	Analysis	20	18	2.00
4	صنف	Analysis	2	3	-1.00
5	ولد	Analysis	1	1	0.00
6	أكتب	application	323	319	4.00
7	حول	application	29	28	1.00
8	حدد	application	20	20	0.00
10	أحسب	application	27	27	0.00
11	أوجد	application	26	27	-1.00
12	أجري	application	3	3	0.00
13	أرسم	application	34	31	3.00
14	أعد	application	8	8	0.00
15	أطبع	application	2	2	0.00
16	اثبت	application	1	2	-1.00
17	بسط	application	2	3	-1.00
18	حل	application	2	3	-1.00
19	ترجم	application	1	0	1.00
20	غير	application	1	3	-2.00
21	نفذ	application	1	0	1.00
22	ماهو / ماهي	comprehension	253	241	12.00
23	كيف	comprehension	31	33	-2.00
24	تحدث	comprehension	26	28	-2.00
25	وضح	comprehension	83	86	-3.00
26	ماذا/لماذا	comprehension	40	37	3.00
27	أشرح	comprehension	9	9	0.00
28	كم	comprehension	8	10	-2.00
29	أعط /هات مثالا	comprehension	8	9	-1.00
30	هل	comprehension	15	17	-2.00
31	صف	comprehension	5	5	0.00
32	علل	comprehension	5	5	0.00
33	أين	comprehension	2	3	-1.00
34	أضف	comprehension	1	1	0.00
35	أوصف	comprehension	2	1	1.00
36	احذف	comprehension	2	1	1.00
37	عبر	comprehension	7	7	0.00

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

No	Arabic Verb	cognitive Level	Expert Evaluation	Electronic Evaluation	The deference
38	أي	comprehension	11	13	-2.00
39	ناقش	Evaluation	7	8	-1.00
40	حاكي	Evaluation	5	5	0.00
41	تحقق	Evaluation	3	3	0.00
42	صحح	Evaluation	1	1	0.00
43	ميز	Evaluation	1	1	0.00
44	كون	Evaluation	8	8	0.00
45	عرف	knowledge	168	168	0.00
46	أذكر	knowledge	171	172	-1.00
47	أختر	knowledge	351	349	2.00
48	ضع	knowledge	335	338	-3.00
49	أكمل	knowledge	60	63	-3.00
50	وصل	knowledge	14	14	0.00
51	عدد	knowledge	6	8	-2.00
52	رتب	knowledge	11	11	0.00
53	مما	knowledge	1	0	1.00
54	أملا	knowledge	80	80	0.00
55	من	knowledge	7	7	0.00
56	استخرج	knowledge	4	4	0.00
57	متى	knowledge	3	3	0.00
58	أنشئ	synthesis	4	3	1.00
59	صمم	synthesis	21	20	1.00
60	طور	synthesis	2	3	-1.00
61	استخدم	synthesis	10	12	-2.00
62	ابني	synthesis	4	3	1.00
63	طبق	synthesis	2	3	-1.00
64	طابق	synthesis	1	1	0.00

TABLE 3 COMPARISON OF BLOOM'S COGNITIVE LEVELS DISTRIBUTION IN THE FINAL EXAMS QUESTIONS.

Cognitive Dimension Level	Experts Evaluation	Electronic Evaluation	The deference between Electronic & Experts Evaluation	Percentage
knowledge	1211	1217	6	0.49%
comprehension	508	506	-2	-0.40%
application	480	476	-4	-0.84%
Analysis	50	48	-2	-4.17%
synthesis	44	45	1	2.22%
Evaluation	25	26	1	3.85%
Sum	2318	2318		

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12806

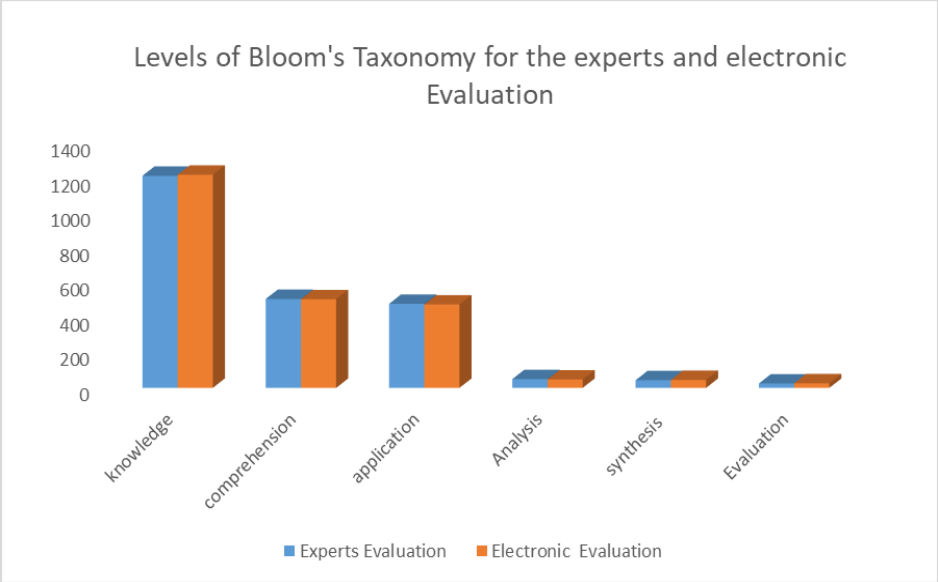


Fig. 5 Comparison of Bloom's Taxonomy for the experts and electronic evaluation Levels of Bloom's Taxonomy for the experts and the electronic evaluation.

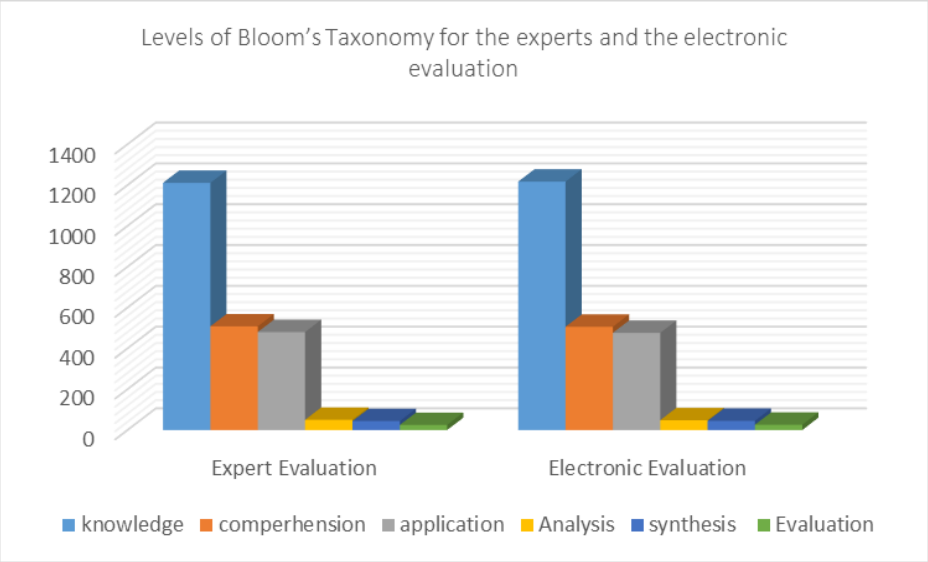


Fig. 6 Level of Bloom's Taxonomy for the experts and electronic evaluation

Results of classification accuracy for expert classification and the system classification:

The findings above showed that the electronic method is significant in classifying questions from multiple domains based on Bloom's taxonomy.

Table 2 shows all the verbs that were used in the exams in the period in which this research was conducted, which is from 2015 to 2018. The verbs used were about 90 verbs that were divided between the six levels of Bloom, but the remembering level has got a large percentage from these verbs, as showed by the analysis, followed by the level of comprehension, and the level of application is in the third place, and according to some divisions of Bloom's levels, these three levels are the lowest level, while the levels of analysis, synthesis and evaluation are called the highest level of thinking.

DOI: 10.17148/IJARCCE.2023.12806

A total of hundred and five (105) final examination papers were reviewed in this study, contained 2318 in which a statistical comparison was done and the result in Table 2 above. It clear showed that there is no statistically significant differences were found in the samples of the study between the results of the analysis in the knowledge level from expert assessment and electronic assessment, the percentage in the deference is 0.49%

Also there are no statistically significant differences were found in the samples of the study between the results of the analysis in the level of comprehension from expert assessment and electronic assessment the percentage in the deference is 0.40%

Also there are no statistically significant differences were found in the samples of the study between the results of the analysis in the application level of expert assessment and the electronic evaluation, the percentage in the deference is 0.84%

Also there are no statistically significant differences were found in the samples of the study between the results of the analysis at the level of analysis from expert assessment and electronic evaluation, the percentage in the deference is 4.17% still is less than 5%.

Also There are no statistically significant differences were found in the samples of the study between the results of the analysis in the level of synthesis from expert assessment and electronic evaluation, the percentage in the deference is 2.22%.

There are no statistically significant differences were found in the samples of the study between the results of the analysis in the evaluation level of expert assessment and electronic evaluation, the percentage in the deference is 3.85%.

IV. CONCLUSION

In this paper we have introduced the problem of classifying final exam questions and learning outcomes in Bloom's Taxonomy. To classify the level of questions we used the automated system methods, using some artificial intelligence methods to classify questions and remove unimportant extras from the questions in order to reach the correct measurement and knowledge level of the question in the cognitive domain.

We have conducted the analysis of the action verbs used in the final exams questions and in the course learning outcomes and a comparison was made with the automated system based on artificial intelligence methods by using programs to extract action verbs from questions and compare them with the list of action verbs as in the taxonomy proposed by Benjamin Bloom.

In conclusion, this will enhance classifying cognitive verbs. However, these verbs are used and cited by academics for writing CLOs and classifying their questions based on blooms taxonomy as this work helps provide more accurate verbs, which in turn will provide more accurate intentional mental skills.

Construction of final exams questions came up as one of the challenges that had an impact on the cognitive levels of questions. If we can gradually modify our way of teaching and questioning towards higher-level cognitive skills according to Bloom's Taxonomy and use them to design exam questions and analyze results, it will greatly improve the quality of assessment in education.

REFERENCES

[1]. Gronlund, N. E. (1968). Constructing achievement tests. Englewood Cliffs, NJ: PrenticeHall.
[2]. Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hili, W. H., & Krathwohl, D. R. Taxonomy of educational objectives: Handbook I: Cognitive domain. New York: David McKay, 1956.
[3]. Duwairi, R.M. (2007). Arabic text categorization. Int. Arab J. Inf. Technol., 4(2), 125-132.
[4]. M. Syiam, Z. Fayed, and M. Habib, "An intelligent system for Arabic text categorization," Int. Journal of Intelligent Computing and Information Sciences, vol. 6(1), pp. 1-19, 2006.
[5]. Khorsheed, M.S. and Al-Thubaity, A.O. Comparative Evaluation of Text Classification Techniques Using a Large Diverse Arabic Dataset. Language Resources and Evaluation, 513-538. (2013) https://doi.org/10.1007/s10579-013-9221-8.
[6]. Ehsan Othman, Ayoub Al-Hamadi, Automatic Arabic Document Classification Based on the HRWiTD Algorithm. Journal of Software Engineering and Applications Vol.11 No.4, April 28, 2018.

DOI: 10.17148/IJARCCE.2023.12806

[7]. Chen Qiao & Xiao Hu, "Text Classification for Cognitive Domains: A Case Using Lexical, syntax and Semantic Features, Journal of Information Science" , 2018, Available at: https://mc.manuscriptcentral.com/infosci

[8]. Harrison J., et, el, "Question Classification According To Bloom's Revised Taxonomy" , 2017, An electronic version of this thesis is available at http://repository.tudelft.nl/.

[9]. Selvia et al, "Automatic Indonesia's Questions Classification Based On Bloom's Taxonomy Using Natural Language Processing" , International Conference on Information Technology Systems and Innovation (ICITSI) Bandung – Bali, November 16 – 19, 2015.

[10]. Dhuha ABDULJABBAR et, al. , "Exam Questions Classification Based On.

[11]. Bloom's Taxonomy Cognitive Level Using Classifiers Combination " , Journal of Theoretical and Applied Information Technology , Vol.78. No.3, 2015.

[12]. Tejalal Choudhary ET AL. , " Improving Teaching – Learning Process using Bloom's Taxonomy and Correlation Analysis" , International Journal of Engineering Research & Technology (IJERT) Vol.3 Issue 6 , 2014 .

[13]. Addin et. al, Automatic Classification of Questions into Bloom's Cognitive Levels Using Support Vector Machines, The International Arab Conference on Information Technology, Naif Arab University for Security Science (NAUSS), Riyadh, Saudi Arabia, 2011.

[14]. El-Khair, I.A., Effects of stop words elimination for Arabic information retrieval: a comparative study. International Journal of Computing & Information Sciences, 2006. 4(3): p. 119-133.

BIOGRAPHY

Dr. Omar Balola was born in Sudan, He is an assistant professor in the Computer Engineerin Department at the **University of Bright Star, El-Brega, Libya.** He held a PhD and M.S in Computer Science from Sudan University of Science and Technology, Sudan in 2017, 2011 respectively. He got B.Sc. (Computer Science) from Omdurman Islamic University, Sudan in 2001. In 2004, he joined the Department of Computer Science, Omdurman Islamic University as a Teaching Assistant, in 2011 became a Lecturer and assistant professor in 2017. At present he teaches courses in Advanced Database, Visual Programming Languages, AI and Intelligent Systems. His current research interests include pattern recognition, image processing, neural network and fuzzy logic.