

**THE MODEL OF EXAMINATION QUESTION  
CLASSIFICATION BASED ON BLOOM'S  
TAXONOMY USING SEMANTIC SIMILARITY  
TECHNIQUE**

**GOH THING THING**

**DOCTOR OF PHILOSOPHY  
(COMPUTER SCIENCE)**

**UNIVERSITI PERTAHANAN NASIONAL  
MALAYSIA**

**2024**

**THE MODEL OF EXAMINATION QUESTION CLASSIFICATION BASED  
ON BLOOM'S TAXONOMY USING SEMANTIC SIMILARITY  
TECHNIQUE**

**GOH THING THING**

This thesis has been submitted to the Centre for Graduate Studies, Universiti  
Pertahanan Nasional Malaysia in fulfilment of the requirements for the Doctor of  
Philosophy (Computer Science).

**January 2024**

## **ABSTRACT**

Bloom's Taxonomy (BT) has generally been used as a guideline in designing a holistic set of examination questions that comprise various cognitive levels. It has been emphasised by Engineering Accreditation Council Malaysia (EAC) and Malaysian Qualifications Agency (MQA) to regulate the quality and standard of education provided by setting the assessment questions aligned with the Course Learning Outcomes (CLO). However, there are inconsistencies in the classification of final examination questions based on Bloom's Taxonomy. This is because it is manually conducted by academics and is susceptible to discrepancies in the understanding of BT among academics. Most of the research work focused on single-sentence questions that were not based on real examination questions. While previous research has explored examination question classification using a semantic approach, it encountered challenges in achieving high accuracy, which is greater than 80%. Therefore, this research aims to introduce a model to perform examination question classification based on BT using a semantic approach with real examination question and striving to attain an accuracy exceeding 80%. A Question Classification Model (QCM) was developed in this research using Natural Language Processing (NLP) approaches, such as the Stanford POS (Part-Of-Speech) tagger, to preprocess the examination questions into word tokens. Subsequently, Stanford Parser Universal Dependencies (UD) was used to identify the important verbs in the examination questions that reflect the thinking action. This was followed by a comparison between the identified verbs and the list of BT verbs using the WordNet Similarity approach. Moreover, this research has studied, evaluated and enhanced each approach to achieve the best performance for the QCM. Overall, the developed QCM achieved a recorded

accuracy rate of 83% in the classification of a set of 200 examination questions based on BT. This research helps to control the assessment quality to meet the classification and fulfil the requirements of Outcome-Based Education (OBE) standards.

## **ABSTRAK**

Klasifikasi soalan berdasarkan Taksonomi Bloom (BT) telah diterima secara meluas dan digunakan sebagai garis panduan dalam penyediaan soalan peperiksaan holistik yang terdiri daripada pelbagai domain kognitif. Pentingnya penggunaan BT telah ditekankan oleh Majlis Pengiktirafan Kejuruteraan Malaysia (EAC) dan Agensi Kelayakan Malaysia (MQA) dalam mengawal standard dan kualiti pendidikan dengan menentukan soalan penilaian yang selaras dengan Hasil Pembelajaran Kursus (CLO). Walau bagaimanapun, terdapat banyak percanggahan pendapat dalam pengelasan soalan berdasarkan BT sekiranya pengelasan dilakukan secara manual oleh staf akademik. Selain itu, kebanyakan kerja penyelidikan tertumpu kepada soalan tunggal yang tidak berasaskan soalan peperiksaan yang sebenarnya. Sementara itu, penyelidikan terdahulu telah meneroka pengelasan soalan peperiksaan menggunakan pendekatan kesamaan. Tetapi, ianya menghadapi cabaran dalam mencapai kejituan yang tinggi, iaitu melebihi 80%. Oleh itu, penyelidikan ini bertujuan untuk memperkenalkan model kerja pengelasan soalan peperiksaan berdasarkan BT menggunakan pendekatan semantik yang menggunakan soalan peperiksaan sebenar, di samping mencapai kejituan melebihi 80%. Model Kerja Pengelasan Soalan (QCM) telah dibangunkan dengan menggunakan pendekatan Pemprosesan Bahasa Semula Jadi (NLP), seperti pemberian tanda Stanford POS (Part-Of-Speech), untuk memproses soalan peperiksaan kepada token perkataan. Seterusnya, Stanford Parser Universal Dependencies (UD) digunakan untuk mengenal pasti kata kerja penting dalam soalan peperiksaan yang mencerminkan tindakan berfikir. Ini diikuti dengan perbandingan antara kata kerja yang dikenal pasti daripada soalan peperiksaan dan senarai kata kerja BT menggunakan pendekatan

Kesamaan WordNet. Selain itu, penyelidikan ini mengkaji, menilai dan memperbaiki setiap pendekatan untuk mencapai prestasi terbaik untuk QCM. Secara keseluruhannya, QCM yang dibangunkan mencapai kadar kejituan sebanyak 83% dalam pengelasan set soalan peperiksaan berdasarkan BT. Kajian ini membantu mengawal kualiti penilaian untuk memenuhi pengelasan dan mematuhi keperluan piawaian Pendidikan Berasaskan Hasil (OBE).

## **ACKNOWLEDGEMENTS**

This dissertation is born out of a journey that began in 2017. This journey would not have been possible without the support of my family, supervisors and friends.

First of all, I would like to thank my family for the full support given all these years. I am especially grateful to my hubby, who supported me emotionally and financially.

I would like to express my gratitude to Dr Nor Azliana Akmal binti Jamaludin, my esteemed supervisor, for all the guidance, support and motivation she provided to me throughout my doctoral studies. I could not have imagined having a better advisor on my journey.

My gratitude also goes to Dr Hassan bin Mohamed and Prof. Dr Mohd Nazri bin Ismail for their insightful comments and encouragement in helping me succeed in my studies. I am also grateful to Prof. Dr Aidy Ali, support staff from UPNM and my friends. Without their support over the past few years, I would not have been able to finish my studies.

Last but not least, I am extending my gratitude to the Faculty of Defence Science and Technology, UPNM for the funding opportunity to undertake my studies.

## **APPROVAL**

The Examination Committee has met on 4<sup>th</sup> October 2023 to conduct the final examination of Goh Thing Thing on her degree thesis entitled “The Model of Examination Question Classification based on Bloom’s Taxonomy using Semantic Similarity Technique”.

The committee recommends that the student is awarded the Doctor of Philosophy (Computer Science).

Members of the Examination Committee are as follows:

**PROF. TS. DR. OMAR BIN ZAKARIA**

Faculty of Defence Science and Technology  
Universiti Pertahanan Nasional Malaysia  
(Chairman)

**PROF. MADYA DR. MOHD RIZAL BIN MOHD ISA**

Faculty of Defence Science and Technology  
Universiti Pertahanan Nasional Malaysia  
(Internal Examiner)

**PROF. DR. MOHD NORDIN BIN ABDUL RAHMAN**

Fakulti Informatik dan Komputeran  
Universiti Sultan Zainal Abidin  
(External Examiner)

**PROF. MADYA DR. SITI SOPHIAYATI BINTI YUHANIZ**

Razak Faculty of Technology and Informatics  
Universiti Teknologi Malaysia  
(External Examiner)

## **APPROVAL**

This thesis was submitted to the Senate of Universiti Pertahanan Nasional Malaysia and has been accepted as fulfilment of the requirements for the degree of the **Doctor of Philosophy (Computer Science)**. The members of the Supervisory Committee were as follows.

**Dr. Nor Azliana Akmal binti Jamaludin**

Fakulti Sains Komputer dan Teknologi  
Universiti Pertahanan Nasional Malaysia  
(Main Supervisor)

**Dr. Hassan bin Mohamed**

Fakulti Sains Komputer dan Teknologi  
Universiti Pertahanan Nasional Malaysia  
(Co-supervisor)

**Prof. Dr. Mohd Nazri bin Ismail**

Fakulti Sains Komputer dan Teknologi  
Universiti Pertahanan Nasional Malaysia  
(Co-supervisor)

**UNIVERSITI PERTAHANAN NASIONAL MALAYSIA**

**DECLARATION OF THESIS**

Student's full name : GOH THING THING  
Date of birth : 29 JULY 1976  
Title : THE MODEL OF EXAMINATION QUESTION  
CLASSIFICATION BASED ON BLOOM'S  
TAXONOMY USING SEMANTIC SIMILARITY  
TECHNIQUE  
Academic session : 2017/2018

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

I further declare that this thesis is classified as:

- ☐ **CONFIDENTIAL** (Contains confidential information under the official Secret Act 1972)\*
- ☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)\*
- ☐ **OPEN ACCESS** I agree that my thesis to be published as online open access (full text)

I acknowledge that Universiti Pertahanan Nasional Malaysia reserves the right as follows.

1. The thesis is the property of Universiti Pertahanan Nasional Malaysia.
2. The library of Universiti Pertahanan Nasional Malaysia has the right to make copies for the purpose of research only.
3. The library has the right to make copies of the thesis for academic exchange.

\_\_\_\_\_  
Signature

\_\_\_\_\_  
\*\*Signature of Supervisor/Dean of CGS/

\_\_\_\_\_  
IC/Passport No.  
Date:

\_\_\_\_\_  
\*\*Name of Supervisor/Dean of CGS/  
Date:

Note: \*If the thesis is CONFIDENTIAL OR RESTRICTED, please attach the letter from the organisation stating the period and reasons for confidentiality and restriction.

\*\* Witness

## TABLE OF CONTENTS

	<b>Page</b>
<b>ABSTRACT</b>	ii
<b>ABSTRAK</b>	iv
<b>ACKNOWLEDGEMENTS</b>	vi
<b>APPROVAL</b>	vii
<b>DECLARATION</b>	ix
<b>TABLE OF CONTENTS</b>	x
<b>LIST OF TABLES</b>	xiii
<b>LIST OF FIGURES</b>	xvi
<b>LIST OF ABBREVIATIONS</b>	xviii
 <b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	1
1.1 Background	1
1.2 Problem Statement	7
1.3 Objectives	8
1.4 Research Scope	9
1.5 Significance of Research	9
1.6 Thesis Outline	10
1.7 Conclusion	11
 <b>2 LITERATURE REVIEW</b>	 12
2.1 Overall View of a Question Classification Model	13
2.2 Education Taxonomy	16
2.3 Natural Language Processing (NLP)	19
2.4 Approaches in Examination Question Classification	22
2.4.1 Rule-based Classifier	24
2.4.2 Learning-based Classifier	26
2.4.3 Statistical-based Classifier	29
2.5 Implementation of Semantic Approach	31
2.6 Comparison of Approaches used in Examination Question Classification	35
2.7 Text Pre-Processing	39
2.8 Semantic Similarity	44
2.9 WordNet	47
2.10 Wu & Palmer Similarity	49
2.11 Research Gap	51
2.12 Proposed Model of Question Classification	53
2.13 Conclusion	56

<b>3</b>	<b>RESEARCH METHODOLOGY</b>	<b>58</b>
3.1	Research Structure	59
3.2	Examination Question Dataset Collection	62
3.3	Overall View of a Question Classification Model	67
3.4	Finalised Question Classification Model (QCM)	69
3.4.1	Question Pre-Processing (QP) module	69
3.4.2	Verb Extraction (VE) module	71
3.4.3	Question Classification (QC) module	74
3.5	Question Pre-Processing Module – Tokenisation	79
3.6	Question Pre-Processing Module – Part-Of-Speech (POS) Tagging	80
3.7	Question Pre-Processing Module – Stop-Words and Punctuations Removal	82
3.8	Verbs Extraction Module – Verb Extraction by Verb Forms	83
3.9	Verbs Extraction Module – Verb and Keywords Identification	86
3.10	Question Classification Module – WordNet Synsets and Similarity	93
3.11	Question Classification Module – Similarity Matrix	95
3.12	Question Classification Module – Root Word Exact Matching Parser Verbs	96
3.13	Question Classification Module – Root Word and Keywords Similarity to Bloom’s Verbs	99
3.14	Question Classification Module – Weightage	103
3.15	Question Classification Module – Bloom’s Level Classification by Tier	107
3.16	Multiple-Sentence Type of Question Classification	113
3.17	Conclusion	115
<b>4</b>	<b>RESULTS AND DISCUSSION</b>	<b>117</b>
4.1	Tokeniser Analysis and Selection	118
4.2	Part-Of-Speech (POS) Tagger Analysis and Selection	119
4.3	Verb Extraction by Verb Forms Analysis and Rules Applied	123
4.4	Root Word and Keywords Identification Analysis, and Rules Applied	128
4.5	Similarity Matrix Measures into Single Score Analysis and Selection	131
4.6	Similarity Matrix Scores Analysis and Sorting	134
4.7	Bloom’s Level Classification by Tier and Similarity Top Score Pick Analysis and Selection	142

4.8	WordNet – Wu and Palmer Similarity Approach and Comparison	144
4.9	Evaluation Result Summary	145
4.10	Question Classification Model (QCM) Result	147
4.11	Conclusion	149
<b>5</b>	<b>CONCLUSION</b>	<b>154</b>
5.1	Suitable Approach for Examination Classification to the BT	154
5.2	Correct Keywords for Question Classification	155
5.3	Validation of the Semantic Similarity Approach for Examination Questions Classification based on the BT	157
5.4	Research Contribution	160
5.5	Future Work	161
	<b>REFERENCES</b>	<b>163</b>
	<b>APPENDICES</b>	<b>170</b>
A.	Collection of Exam Question	170
	<b>BIODATA OF STUDENT</b>	<b>183</b>
	<b>LIST OF PUBLICATIONS</b>	<b>184</b>

## LIST OF TABLES

<b>TABLE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
2.1	Anderson's Revisions on Bloom's Taxonomy verbs (Jayakodi et al., 2015)	18
2.2	NLP Main Focused Areas	20
2.3	Summary of NLP Approach – Keyword Matching and Rule-Based	36
2.4	Summary of NLP Approach – Machine Learning and Semantic	37
2.5	Comparison between different Studies on Questions Classification	38
3.1	Statistic of the dataset of 200 examination questions in subject domain, Bloom's level and sentences count	63
3.2	Statistic of subjects offered in an Engineering Diploma programme of UOW Malaysia KDU University College	66
3.3	Q1 is single sentence type of question while Q2 is multiple-sentence type of question and its tokenised single sentences - Q2S1 and Q2S2	79
3.4	Word Tokenisation	80
3.5	POS Tagging	81
3.6	The default Stop Words List from NLTK and Punctuations list	82
3.7	Stop Words Removal and Punctuations Removal	83
3.8	Verbs Forms Description and Example	84
3.9	Scenario of No Verb Form with “VB” tag but with other Verb Forms such as “VBP” in a Single Sentence Question	84
3.10	Scenario of Multiple Verbs with “VB” tag in a Single Sentence Question	85
3.11	Stanford Parser Universal Dependencies Tree Transformed into a Table	88
3.12	Finalised Root Word, Keywords Level 1 and Level 2 of Single Sentence without a Verb tagged with “VB”	91
3.13	Finalised Root Word, Keywords Level 1 and Level 2 of Single Sentence with Multiple Verbs	92
3.14	Synsets of the Words – car, automobile, boat and cat	93

## LIST OF TABLES

<b>TABLE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
3.15	Wordnet Similarity Scores	94
3.16	Similarity Matrix stored Similarity Scores between 'car' and 'cat'	96
3.17	Root Word Exact Matching Scenarios	97
3.18	Root Word Exact Matching Bloom's Verbs Scoring	98
3.19	Statistics of Bloom's Verbs per Bloom's Levels and their Synsets	99
3.20	Summarising Three-dimensional WordNet Similarity Matrix into 6 Scores with respective to each Bloom's level	102
3.21	Root Word Exact Matching Bloom's Verbs Scoring After High Weightage Application	104
3.22	Weightage Application on the Root Word and Keywords Similarity to Bloom's Verbs Result	105
3.23	High, Normal and Low Weightage Application on the Root Word and Keywords Similarity to Bloom's Verbs Result	107
3.24	Bloom's Level Classification at Tier 1	109
3.25	Bloom's Level Classification at Tier 2	111
3.26	Bloom's Level Classification at Tier 3	112
3.27	Bloom's Level Classification for Multiple-Sentence Type of Questions	114
3.28	Summary of Research Achievement by Research Objectives	116
4.1	The differences in Word Tokenisation result by the PunktSentence Tokeniser and the CoreNLP Tokeniser	119
4.2	The Grading System for the POS Taggers Performance	120
4.3	The POS Taggers Performance Result	122

## LIST OF TABLES

<b>TABLE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
4.4	The ERR2 errors due to some Verbs used as the First word in a Single-sentence Question, tagged wrongly as Noun	125
4.5	The ERR2 errors due to some Verbs Not tagged as “VB” but tagged as other verb forms – “VBD”, “VBG”, “VBN”, “VBP”, and “VBZ”.	126
4.6	Effectiveness of Verb Extraction Process After Implementation of Rules	127
4.7	The Root Word Extraction from the Stanford Parser Universal Dependencies Tree and its Rules	131
4.8	Bloom’s Level Classification for Multiple-sentence Type of Questions	132
4.9	The Effectiveness of Statistical Measures – Mean, Maximum and Median in the 200 Examination Questions dataset	133
4.10	Similarity Scores Count of a Root Word – “draw” to each Bloom’s levels	135
4.11	Effectiveness of Synsets with Verb tag only in the 200 Examination Questions dataset	135
4.12	The Sorted Similarity Scores of W0 – “draw” by Top 10, 20, 30, ... and its Bloom’s Level Classification based on Highest Score.	140
4.13	The Sorted Similarity Scores of W0 – “write” by Top 10, 20, 30, ... and its Bloom’s Level Classification based on Highest Score	140
4.14	Performance Evaluation Between Wordnet – Wu and Palmer Similarity Approach and Others Using The Same 200 Questions Dataset	145
4.15	Summary of All Evaluation Results	146
4.16	Question Classification Model (QCM) Performance Results by Single and Multiple-sentence Questions	148
4.17	Question Classification Model (QCM) Performance Results by Domains	148
4.18	Summary of Research Achievement by Research Objectives	152

## LIST OF FIGURES

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
1.1	Accreditation feedback from EAC for Bachelor of Mechanical Engineering, UOW Malaysia KDU UC	5
2.1	Flow of Literature Review	12
2.2	Question Classification Model - A Pipeline Processing	14
2.3	Bloom's Taxonomy and its Revision	16
2.4	NLP Approaches in Examination Classification	23
2.5	POS Tagging Example	40
2.6	Penn Treebank Tagsets	40
2.7	The "back" word and its POS tags	41
2.8	Knowledge Graph on Synonyms	45
2.9	WordNet Semantic Network (Mark Sanderson, 1994)	48
2.10	Example of Ontology Extract (Slimani et al., 2008)	51
2.11	Model of Question Classification based on Bloom's Taxonomy using Universal Dependency and WordNet	54
3.1	Research structure to develop a model of Question Classification according to BT	60
3.2	Programme structure of Diploma in Electrical and Electronics Engineering of UOW Malaysia KDU University College	64
3.3	Programme structure of Diploma in Mechanical Engineering of UOW Malaysia KDU University College	65
3.4	Research structure to develop a model of Question Classification according to BT	68
3.5	Detail Flow of the Question Pre-processing (QP) module of the finalised Question Classification Model (QCM)	71
3.6	Detail Flow of the Verb Extraction (VE) module of the finalised Question Classification Model (QCM)	73
3.7	Detail Flow of the Question Classification (QC) module of the finalised Question Classification Model (QCM)	75

## LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
3.8	Detail Flow of the continued Question Classification (QC) module of the finalised Question Classification Model (QCM)	77
3.9	Stanford Parser Tree and Triples	87
3.10	Stanford Parser Tree of a Multiple Verbs in Single Sentence Scenario	90
3.11	The Root Word, Keywords Level 1 and Level 2 Finalisation Flow	91
3.12	Three-dimensional Matrix to store Similarity Scores of W0 – Root Word to W1 – Bloom’s Verbs	100
4.1	Stanford Parser Universal Dependencies Tree with a Noun at Level 0	129
4.2	Stanford Parser Universal Dependencies Tree with a Verb with non “VB” tag at Level 0.	130
4.3	The Unsorted Two-dimensional Similarity Matrix Between the RW – “write” and the BVs of the Bloom’s Level 1	136
4.4	The Sorted Two-dimensional Similarity Matrix Between the RW – “write” and the BVs of the Bloom’s Level 1	137
4.5	The Sorted Similarity Scores Summarised by Mean approach into Sorted Top to Low values	138
4.6	The Single Similarity Scores of the RW – “write” per Bloom’s Levels After Summarising the Similarity using Statistical Mean Approach	139
4.7	Bloom’s Level Classification by Similarity Top Score on the 200 Examination Questions Dataset	141
4.8	Bloom’s Level Classification by Tier Result Analysis	143
4.9	Bloom’s Level Classification by Tier with different Top Similarity Score Pick	144

## LIST OF ABBREVIATIONS

UPNM	Universiti Pertahanan Nasional Malaysia
BT	Bloom's Taxonomy
BV	Bloom's Verb
CLO	Course Learning Outcomes
ERR1	Error 1
ERR2	Error 2
ML	Machine Learning
ME	Maximum Entropy
NLP	Natural Language Processing
RW	Root Word
JCN	Jiang, Conrath measure
KW	Keyword
LCH	Leacock and Chodorow
left3words	"english-left3words-distsim" model
LIN	Lin measure
POS	Part-Of-Speech
PLO	Programme Learning Outcomes
QC	Question Classification
QCM	Question Classification Model
QP	Question Pre-processing
SVM	Support Vector Machine
TP	True Positive
UD	Stanford Parser Universal Dependencies
VE	Verbs Extraction
WUP	Wu and Palmer's Similarity

## **CHAPTER 1**

### **INTRODUCTION**

This chapter introduces Outcome-based Education which focuses on the outcomes that are commonly classified based on Bloom's Taxonomy (BT). This is followed by the examination questions which are a common tool used to access the outcomes. Thus, the examination questions should be classified based on BT and in sequence to align with the outcome classification also. However, various discrepancies occur in question classification due to the inconsistency of questions in BT.

#### **1.1 Problem Background**

Outcome-based Education (OBE) is an education theory that focuses on outcomes – what students should learn and their ability to apply the lesson after the learning process (Qadir et al., 2020). To ensure that the students achieve the defined outcomes, Malaysian Qualifications Agency (MQA) has implemented a Malaysia Qualifications Framework (MQF) to regulate the quality and standard of higher education provided. This model is defined as an instrument in developing and

classifying qualifications based on a set of criteria that have been nationally agreed upon since 2008.

Five learning outcome clusters have been defined and used as a qualification guideline for all Technical and Vocational Education and Training (TVET) to obtain their programmes accreditation by MQA. Each programme is structured for a specified duration and learning volume to achieve the stated Programme Learning Outcomes (PLO). These factors contribute to an award of a qualification, such as a diploma certification, a Bachelor's degree, or a Master's degree among others.

The PLOs are the specified knowledge, skills, attitude, and abilities to be acquired and demonstrated by students upon graduation (MQA, 2017). All PLOs should be aligned to the learning outcome clusters set by MQA. The achievement of PLOs is based on the Course Learning Outcomes (CLOs) of each module in the programme. The achievement of CLOs is heavily dependent on the score obtained in each assessment component.

The assessment components are the final examination, tests, projects and assignments distributed to students within a semester. To achieve the CLOs followed by PLOs, the assessment components should be designed in line with the CLOs set for the modules. Furthermore, the questions set in each assessment component plays an important role in assisting students in attaining the targeted CLOs where each CLO is set with different thinking order based on Bloom's Taxonomy (BT). Specifically, lower-order thinking questions emphasise foundation skills and practices, while

higher-order thinking questions involve more complex thinking and creative problem-solving.

Established by Benjamin Bloom's during the 1950s, BT is a method of categorising the levels of reasoning skills required in classroom situations. Six levels in the BT are present, with each requiring a higher level of abstraction from the students. Notably, BT has been widely used as a guideline in designing a holistic examination question, which consists of various cognitive levels (Omar et al., 2012). Teachers are recommended to move students up the BT levels as they progress in their knowledge (Forehand, 2010).

The examination questions classification shall be aligned with the CLO which was defined during the curriculum design. Each CLO is specified with an action verb that corresponds to a Bloom's Taxonomy (BT) verb. Example of a course learning outcome, "Describe concept, principles and theories relating to area of physics and engineering". The action verb "Describe" aligned with the BT verbs from level 2 which is "Understanding". Consequently, academician shall design the examination question which is able to reveal the achievement of the respective course learning outcome.

The current practice in most universities, academicians manually categorise the examination questions for each assessment into the BT levels based on their understanding of BT levels. The question classification was done manually which may vary from one academician to another (Yusof & Chai, 2010; Sulaiman et al., 2020; Jayakodi et al., 2016). Academicians with a clear understanding of the

cognitive levels of BT are able to create sets of examination questions in line with the targeted CLOs. However, most academicians are not familiar with the proper implementation of the cognitive levels of BT (Contreras et al., 2021; Kumara et al., 2019).

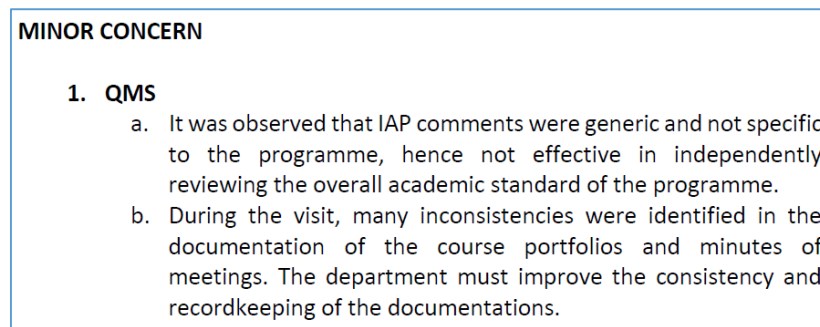
When formulating examination questions, some of the academicians do not follow the prescribed list of BT verbs as a reference. For example, a question such as “Determine the mass and weight of the air contained in a room with dimensions of 6m x 6m x 8m.” has the action verb “Determine” that is not in the BT verb list. Therefore, academicians may classify such questions into Bloom’s levels based on their individual interpretation, leading to the inconsistency in the classification process.

In another case, the question “Calculate the potential difference between points a and b in Figure 3.” was supposed to classify to third level of BT, “Applying”, given that it necessitates students to perform calculations for the potential difference, however, it was classified to the fourth level of BT, “Analysing”. The inconsistency in classification was the result of the academician’s interpretation, which expected students to engage in an analytical process by referring to the provided diagram.

In addition to the misclassification due to lack of understanding of OBE implementation, ambiguity in BT verbs can also contribute to misclassification. The classification of questions based on BT is commonly based on the verbs used in the examination questions. The verbs are extracted from the examination questions and mapped to the list of verbs in BT. However, some of Bloom’s verbs are ambiguous

when the verbs fall into more than one level of BT (Osadi et al., 2017; Das et al., 2020; Shaikh et al., 2021). Therefore, issues are present in the BT classification of examination questions with these verbs that caused inconsistent BT classification by different academicians in the same examination questions.

As a result of the misclassification, concern was raised in the accreditation summary report 2022 for Bachelor of Mechanical Engineering, UOW Malaysia KDU University College by Engineering Accreditation Council (EAC), as shown in Figure 1.1.



**Figure 1. 1 Accreditation Feedback from EAC for Bachelor of Mechanical Engineering, UOW Malaysia KDU UC**

In Malaysia's Institutions of Higher Learning (IHL), the common practice involves appointing one or two academicians as moderators during the examination question moderation process to reduce the inconsistency in question classification. The moderators are responsible to verify the alignment of the examination questions with the CLOs for the respective modules. Feedback is provided to the module lecturer, who is then required to revise the examination questions to ensure all examination questions align with the CLOs. After the revision, the examination paper undergoes a second round of review by a different moderator. This process has

effectively minimised the misclassification of examination questions, however, it is time-consuming.

In recent years, there has been an increase of interest among the researchers in automating the examination question classification based on BT cognitive domain. Various techniques have been used such as syntactic features, which focused on the structure and the pattern of the question and semantic features which consider the meaning of significant verbs identified from the question. Natural Language Processing (NLP) was commonly used to identify the action verb from the examination question. Machine learning or semantic similarity was used to classify the examination question. Some researchers (Sangodiah et al., 2017; Sulaiman et al., 2020; Contreras et al., 2021) have employed machine learning and semantic similarity (Jayakodi et al., 2016; Diab & Sartawi, 2017) to classify questions, but machine learning requires large question datasets for accurate results. This led researchers to suggest using a semantic approach for better outcomes. However, the semantic approach alone has not achieved a good result. Thus, researchers continue to use both machine learning and semantic approaches in their research (Mohammed & Omar, 2020; Mohamed et al., 2019).

Gathering a substantial set of examination questions can be challenging and most research studies have not used the real-exam question. They mostly focused on single-sentence question (Jayakodi et al., 2016; Sangodiah et al., 2017; Mohammed & Omar, 2018). However, in practice, exam questions are often longer and more complex, particularly those involving case studies or scenarios with higher thinking order.