

Assessing the end-of-Semester Examination Papers During the Implementation of The Bologna Process: Bloom's Taxonomy as a Framework

Sarkawt M. Qadir¹, Rukhsar M. Omar², Muhammad H. Rasheed³, Chachan J. Mohammed⁴

¹Department of English Language, College of Education, University of Raparin, Ranya, Kurdistan Region, Iraq.

²Kindergarten Department, College of Basic Education, University of Raparin, Ranya, Kurdistan Region, Iraq.

³Department of Chemistry, College of Science, University of Raparin, Ranya, Kurdistan, Region, Iraq.

⁴Department of Special Education, College of Basic Education, University of Duhok, Duhok, Kurdistan Region, Iraq.

ABSTRACT

One of the significant transformations in the Kurdistan Region of Iraq (KRI) is the implementation of the Bologna process. This alteration of the traditional system to the Bologna process system has also impacted the assessment schemes. More weight has been given to formative assessment that can help improve students' grades and ease success in the courses. Nonetheless, the final exam still carries most of the grades. Therefore, setting appropriate questions that can meet all the cognition levels represented in Bloom's taxonomy assists in raising students' cognition to higher levels rather than only assessing bookish knowledge that is located on the baseline of Bloom's taxonomy. To this intent, the present study endeavored to identify if instructors at colleges of Nursing and Science adhere to the various dimensions of Bloom's taxonomy that are central to the Bologna process. More importantly, the association of each of the demographic variables to the level of the final examination questions was also examined. The study employed a quantitative method to tackle the topic. Totally, a sample of 75 final examination papers was collected from instructors, including 524 individual questions. The findings revealed that most of the examination papers revolved around low-order thinking questions and the association of the rate of success to the level of the questions was highly statistically significant.

KEYWORDS: Assessment, end-of-semester examination paper, Bloom's Taxonomy, Bologna Process, rate of success.

1. INTRODUCTION:

Evaluation, at its heart, is an essential tool in decision-making and generally refers to determining the effectiveness and value of something, more specifically a program or a course of study at schools or a higher education institution. Tests and exams are integral parts of evaluation that are carried out to measure students'

level of knowledge or performance, their achievement, and the progress that has been reached by the students after they go through the teaching/learning process (Abduh, 2015; Himmah et al., 2019; Richards and Schmidt, 2010). Having obtained information regarding students' knowledge, achievement, and course outcomes, it can be easier for educational institutions to make more rational decisions (Hughes, 1989). Since evaluation has this significant status in achieving educational goals, setting appropriate questions for exams that can satisfy diversified cognitive levels proves a determinant factor in performing an evaluation, particularly at the university level (Mohammed and Omar, 2020; Naranayan and Adithan, 2015; Swart, 2010).

In the education process, teachers require to administer testing to measure the teaching objectives they aim to achieve. This presupposes that teaching and testing are closely intertwined and the success of one

Koya University Journal of Humanities and Social Sciences (KUJHSS), Volume 6, Issue 1, 2023.

Received 1 Apr 2023; Accepted 4 May 2023,

Regular research paper: Published 20 Jul 2023

Corresponding author's e-mail: sarkawt.en@uor.edu.krd

Copyright ©2023. Sarkawt M. Qadir, Rukhsar M. Omar, Muhammad H. Rasheed, Chachan J. Mohammed, this is an open access article distributed under the Creative Commons Attribution License.



leans on the lucrative application of the other. Put another way, the quality of a test paper usually determines the quality of the teaching process and materials and vice-versa (Abduh, 2015). Therefore, designing questions that suit students' levels and correspond to the instruction applied in class is deemed significant and can reconcile the interconnection between teaching and testing that seems to be the status quo of the current educational context. In this regard, instructors' assessment literacy can become a contributing factor to the quality of questions (Leung, 2014).

In addition to the evaluation of students, the evaluation of question papers by experts can have significant effects on the quality of both the teaching/learning process and the questions. At university in the current educational context, test questions are not assessed and no improvements are recommended to check the quality of questions or tests designed for university students. University instructors need to set question papers in a way that is inclusive of higher-order thinking and critical thinking to encourage creativity, and critical reasoning as, based on Demir and Eryaman (2012), one of the principal goals of education is to cultivate students to be capable of creating novel ideas instead of reinventing the wheel relying on factual recall and not educating individuals who accept everything without questioning.

Sundry categorizations have been proposed in the literature for question classification. The simplest and most basic categorization has been made on the basis of complexity and abstraction such as the one proposed by Bloom. For instance, the questions that children ask can significantly differ from those of adults (Swart, 2010). A simple question like, "Does America support Kurdistan independence?" that requires a simple Yes or No differs from a more complex open-ended question, "What is the impact of internal conflicts on Kurdistan's independence?" that requires more complex reasoning and thinking. Bloom's Taxonomy is one such categorization that helps classify questions based on their complexity levels. At the lowest level, students are required to memorize facts or concepts. As the level rises, the level of abstraction and complexity increases. The lower levels of the taxonomy represent the base knowledge while the higher levels represent deeply processed knowledge (Naranayan and Adithan, 2015; Swart, 2010).

What is problematic in the present educational context is that students encounter final exam questions that are virtually based on multiple-choice and true-false items located within the low-order thinking skills of Bloom's Taxonomy that assess students' capacity to recall

bookish knowledge and seem to lack other higher-order reasoning and thinking skills such as evaluation, synthesis, and analysis in both pre-university and university education, as observed. This can occasionally be one reason behind students' successes or failures as it might be associated with the quality, category, and/or difficulty level of the questions. Some colleges, such as the science or medical colleges, are not within the domain of education, and instructors in those colleges are more or less not acquainted with the various methods of evaluation and might lack the knowledge required for designing well-qualified questions, as confirmed by (Naranayan and Adithan, 2015). Therefore, undertaking research in this area in the present cultural and educational context might shed more light on the areas of weaknesses and contribute to the literature available in this area of inquiry.

Thus, the major objective of the present study is to assess and evaluate the quality of final examination questions set for both colleges of Nursing and Science at a public university in KRI during the execution of the Bologna process, applying Bloom's Taxonomy as the framework, which is a model for instruction and assessment. More specifically, it aims to understand if examiners at these two colleges adhere to the various levels of complexity proposed by Bloom in designing and preparing questions and then categorize the questions into low-order and high-order reasoning questions. More precisely, the below research questions guide the progress of the current study:

1. What is the level of the end-of-semester examination questions in both colleges of Nursing and Science based on Bloom's Taxonomy?
2. Does the level or quality of the questions vary according to instructors' teaching experience, education level, and scientific title as well as the success rate of students?

2. THE REVIEW OF THE LITERATURE

2.1 Bloom's Taxonomy

Bloom's Taxonomy, which is also known as the educational objectives taxonomy, is recognized as the most significant and leading educational theory in the arena of education and is currently used in most UK universities as the foundation for both teaching and assessment (Newton et al., 2020). The taxonomy was formulated by Bloom and his collaborators in 1956 for categorizing educational learning objectives (Almerico and Baker, 2004; Newton et al., 2020). It comprises three educational dimensions: the cognitive dimension, the

affective dimension, and the psychomotor domain. The taxonomy tackles learning outcomes that are sourced from a specific lesson or an entire course to increase knowledge (the cognitive domain), promote skills or physical activity (the psychomotor dimension), and enhance emotional aptitude or keep balance (the affective dimension) (Bloom, 1956). The goal of using the taxonomy is to stimulate higher-order reasoning and thinking by promoting lower-level cognitive skills. In other words, the taxonomy is educationally hierarchical; this implies that learning at each high level is based on acquiring prerequisite information at a lower level. To further explain, learning to apply (the application level) facts is based on learning to understand what the facts mean (the comprehension level) (Shabatura, 2013). The cognitive domain within this categorization primarily

focuses on learners' cognitive level in a written test (Omar et al., 2012) to test overall cognitive levels. It focuses on the students' learning to distinguish between high-order questions and low-order questions used in assessments. Most final exam questions include verbs that play a significant role in recognizing the cognitive quality of the question. For instance, the verbs 'list' and 'state' are closely associated with the level of 'Knowledge' in the cognitive domain (Mohamed et al., 2019). The six levels of Bloom's categorization are displayed in the table below.

Table 1. Bloom's Categorization for the Cognitive Domain (Adapted from Swart, 2020)

Objective	Definition	Synonym	Illustrative Verbs	Level
Evaluation	determining the value of the system	Assessment	Confirm, justify, determine, conclude, analyze, evaluate	Highest
Synthesis	Joining elements to form a system	Combination	design, combine, predict, construct, formulate, propose, improve	High
Analysis	dividing a system into its parts	Breakdown	Distinguish, classify, contrast, categorize, compare	High
Application	The use of learned material in new situations	Use	Demonstrate, change, solve, modify, show, calculate, use	Low
Comprehension	Grasping the meaning of material	Understand	Convert, explain, summarize, rearrange, estimate, derive, review, relate	Low
Knowledge	Recall of certain elements	Information	State, list, describe, define, sketch, identify, insert, complete	Lowest

Note. This table explains how the analysis proceeded and does not include all the verbs used for the analysis.

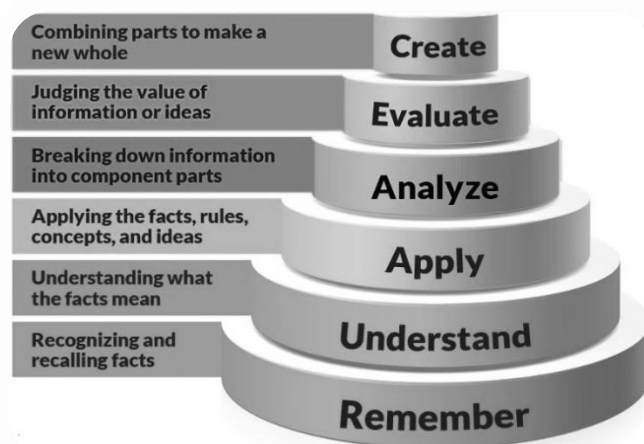
Accordingly, the first lowest level, knowledge, which is employed to determine the students' memory capacity including remembering, is the foundation for all learning and imperative for later higher levels. The second low level, comprehension, involves grasping the meaning of the material. The third level involves the application level that is used to modify a conundrum and evaluate students' application skills. Analysis constitutes a higher-order level of thinking, the fourth, in which students are required to divide the whole system into component parts. Unlike the fourth level, the fifth requires the combination of the various parts into one complete system, called synthesis. The highest level, which is the sixth, is used to evaluate certain techniques and students need to grasp facts (Swart, 2020), as

illustrated in Table 1. Based on Anderson and Krathwohl (2001) and Swart (2010), knowledge and comprehension levels are subsumed under low-order reasoning while evaluation, synthesis, analysis, and application can be subsumed under high-order thinking skills. Nonetheless, based on Pappas et al. (2012) and Yahya et al. (2012), the three levels at the bottom of Table 1 involving knowledge, comprehension, and even application measure low-order thinking skills while the second three levels located at the top of the table involving analysis, synthesis, and evaluation measure high-order thinking skills, according to which the question papers of this study are analyzed and categorized.

Later, the taxonomy was revised by Anderson and Krathwohl (2001) in which only the noun categories changed into verb categories and

evaluation was transposed by creation, as illustrated in Figure 1 below.

Figure 1. The Revised Bloom's Classification for the Cognitive Dimension (Shabatura, 2013)



2.2 Questions and Questioning

To assess students' learning, there are many types of assessments. One of the most prevalent categories involves the use of formal and end-of-year written exams. Although written exams are traditional ways of assessment, they prove a universal method in most current educational institutions worldwide. Intertwined with written exams is a question that is raised to examine the cognitive levels of students (Omar et al., 2012). Questions and questioning have long ago been used as effective instructional and assessment techniques in educational settings (Jiang, 2020). Questioning properly is a sophisticated skill in which many teachers are not proficient (Moore, 2009). Effective teachers need to be proficient and skilled at asking questions at low-level questions and high-level questions (Wiseman and Hunt, 2014). Prior research has shown that questioning proves the second most favored instruction method after lecturing and that instructors spend 35 to 50 percent of their teaching time holding questioning sittings (Moore, 2009). However, what is most relevant here involves the use of questions for the assessment purpose, more specifically the final exam questions.

Designing questions effectively is required for university instructors in particular due to several reasons. Firstly, teacher assessment literacy, i.e., the teacher's capability to design or select superb assessments has been reported to strongly affect

students' achievement (Jiang, 2020). Secondly, appropriately designed questions can help facilitate student learning, maintain student attention, and provide opportunities for rehearsal. They can be utilized to gain information, provoke thinking, deflect reasoning (Crisp et al., 2018; Swart, 2010), motivate students, evaluate students' preparation, develop critical thinking skills, and recognize the achievement of objectives (Crisp et al., 2018; Mohammed and Omar, 2020; Moore, 2009; Swart, 2010).

Questions require to be asked at the appropriate level, must be of the appropriate category, and be correctly worded (Moore, 2009). Sundry categorizations have been proposed. One categorization is based on display questions that require familiar information and referential questions that require unfamiliar information. Another categorization classifies questions into convergent and divergent types. The former concerns low cognitive remembering of already instructed information, while the latter expects differing responses and requires a high level of reasoning. Yet another dominant categorization involves Bloom's Classification which organizes questions from a low level to a high level, that is, from knowledge to evaluation (Jiang, 2020; Richards and Schmidt, 2010).

There is a strong connection between final exam questions and Bloom's Taxonomy. The level of each question should locate within a specific level in the taxonomy. University instructors do not have to include many high-order questions in a first-year final examination paper. Oppositely, they cannot include many low-order questions in a fourth-year final examination paper. The percentage of low-order questions should decrease as students advance to higher academic levels (Omar et al., 2012; Swart, 2020) and vice-versa.

2.3 Related Work

Much work (Crisp et al., 2018; Mohamed et al., 2019; Naranayan and Adithan, 2015; Omar et al, 2012; Swart, 2010) has been devoted to tackling question quality and question classification. Crisp et al. (2018) used the views of question administrators to investigate conceptualizations of question quality by showing them the exam questions and demanding them to assess the level of the questions. The raters defined question quality in terms of simplicity, clarity, and appropriate and consistent use of keyword verbs such as 'define, explain, etc.', and testing the intended

knowledge. Other researchers focused on question classification based on Bloom's Categorization and revealed that most of their questions did not accord to the levels of Bloom's Categorization. Naranayan and Adithan (2015) explored the engineering faculty instructors' familiarity with Bloom's Classification and evaluated final exam question papers considering high-order and low-order thinking skills. They revealed that the engineering instructors had a low level of awareness regarding Bloom's classification of cognitive levels and those who had awareness of the taxonomy chose not to apply it to their teaching and assessment. Additionally, they discovered that the frequency of high-order questions was far lower than the expectation of universities. Swart (2010) distinguished between two sorts of questions, namely high-order and low-order questions employing Bloom's Classification in an engineering program called Electronics. The findings of his investigation displayed that a high proportion of the exam questions involved 'application' questions and academics adopted more low-order questions than high-order questions. He suggested that a balance needs to be struck between these two categories of questions and higher levels of questions should be incorporated in final exam questions in electronics as they might need it in their future careers.

Efforts have also been made to categorize questions using machine learning techniques. One such attempt involves Omar et al. (2012) who adopted a rule-based method through natural language processing for categorizing questions into corresponding Bloom's Classification levels in a programming subject. They found that such an approach successfully assisted in the recognition of question categories into accurate cognitive levels. Additionally, Mohamed et al. (2019) proposed an amalgamation technique of both syntactic and semantic methods to correctly categorize questions into various cognitive levels of Bloom's Classification. They utilized three techniques of machine learning, i.e., classifiers involving Naive Bayes, Support Vector, and J48 with the combination technique and without the combination technique. They discovered that the application of the techniques with the combination technique outperformed their application without the combination method.

A careful examination of the literature reveals that most studies have reported the categorization of exam questions using Bloom's Classification for many fields such as engineering, medicine, and computing. Most of those studies revealed a low level of instructors'

awareness of Bloom's Taxonomy and that their questions were more attributed to low-order than high-order which is essential for developing creative, analytical, and critical thinking skills. However, limited studies have investigated the classification of questions in the nursing and science fields such as chemistry, physics, and biology, particularly in the Kurdish context which, to date, no studies have been reported. Therefore, the present study will utilize Bloom's Classification as a framework to classify final exam questions into the different cognitive levels to identify Kurdish instructors' level of questions in those two colleges that might function as feedback to testing and evaluation in the context of this study. Similar to Mohamed et al. (2019), it will use a combination method, i.e., a method that considers an amalgamation of syntactic and semantic methods to categorize the questions.

3. METHODOLOGY

3.1 Research Design:

The present study analyzed the final examination question papers written by the faculty instructors at both Colleges of Nursing and Science at a public university in KRI during the implementation of the Bologna process. To this intent, the researchers utilized a quantitative research design that can be characterized by certain features: First, a quantitative design creates models and theories displayed in mathematical expressions; second, the researcher's role is neutral to avoid any bias; third, numerical data is collected and analyzed objectively through statistical techniques; fourth, association between variables can be shown quite easily (Salkind, 2010). More importantly, the descriptive approach is a fundamental approach that describes the situation as it avails in its current state. It involves the identification of the properties of a specific phenomenon based on an observational basis or the exploration of the correlation between two or more phenomena (Williams, 2007).

3.2 Data Collection

This study collected data from a corpus of all the final exam question papers, equaling (75) question papers featuring (524) individual questions, set by the faculty instructors at both colleges. The questions belonged to different education levels as well as programs and departments of the College of Nursing and Science. Additionally, the question setters'

scientific titles were also various including assistant lecturer, lecturer, assistant professor, and professor, possessing master's and doctorate degrees. The questions included various categories of questions endorsing definitions, enumeration, labeling,

applying, true-false, short and long answer questions, and multiple-choice questions as well as matching exercises. Table 2 represents samples of questions for each cognitive level categorized based on the key verb found in the corpus.

Table 2. Sample Questions from the Corpus for the Cognitive Levels of Bloom's Classification

Cognitive Level	Question Examples from the Data
Knowledge	- Define the following terms: Rate, activation energy, order of reaction.
Comprehension	- Discuss the Side effects of Chemotherapy briefly?
Application	-Sara is a 35 years-old G6, P4, Ab1, was admitted 12 hours after normal vaginal delivery, she had been having vaginal bleeding, her instable vital signs and chills. What is nursing intervention to her condition?
Analysis	- Calculate neutron separation energy for $^{16}_7\text{N}$ nucleus? $1\text{amu}=1.66 \times 10^{-27} \text{ kg} = 931.48 \text{ MeV}/\text{C}^2$ Mass of $^{15}_7\text{N}= 15.000109 \text{ amu}$, mass of $^{16}_7\text{N}=16.006102 \text{ amu}$ Mass of neutron= $1.6749286 \times 10^{-27}$
Synthesis	- Prepare the following compounds: 1. Hexan-3-ol by Oxymercuration-demercuration start from hex-3-ene. 2. Propoxybutane by Williamson ether synthesis.
Evaluation	- Determine the spin and parity of the ground state of these nuclei $^{15}_7\text{N}$, $^{41}_{20}\text{Ca}$.

Note. The question examples have been directly borrowed from the corpus.

3.3 Data Analysis

To analyze the data, the researchers codified the question papers and inspected the documents for labeling and annotating the questions into the diversified cognitive levels of Bloom's Classification based on the key verbs found in each question as the verb plays a significant role in assessing the level of the questions in Bloom's Categorization (Mohammed and Omar, 2020). Therefore, the rating scale displayed in Table 1 above was applied to determine the robustness of the questions. Subsequently, the questions were categorized into two major levels of analysis involving high-order and low-order questions.

The collected data from the examination question papers were then examined and descriptive statistics such as percentages and frequencies were utilized to indicate the level of the questions. Then, inferential statistics such as regression coefficients were applied to display the impact of the demographic variables (teaching experience, rate of success, academic title, and academic degree or qualification) on the level of the questions. Regression analysis is one of the most widely utilized statistical techniques that can be useful for analyzing multifactor data, proving an

outstanding method for studying functional interrelationships among a certain set of variables and it is better than the other correlations because it gives us a better summary of the connection between two specific or more specific variables (Chatterjee and Hadi, 2006).

3.4 Validity and Reliability

To accurately classify the questions, the present study did not only consider the keywords found in the questions, rather it depended on weight, considering verbs as the most important elements followed by nouns and adjectives (See Omar et al., 2012) as well as WH-question words. More importantly, the syntactic and semantic aspects of the question categorization were considered. Each exam question paper was independently scrutinized and categorized by the researchers. When differences in the question classification had raised, the researchers would discuss them. The researchers belonged to various specialties including social psychology, English language and linguistics, physics, and educational psychology which are quintessential in determining question categories. Both inter-rater reliability (IRR) and intra-rater reliability were calculated for the categorization of the questions

among the raters, obtaining a high degree of (IRR= 82%) and (INRR= 89%) respectively.

It is worthwhile to mention that the researchers strived to accurately determine the classification of the questions by paying strong heed to the keywords and not only the key verbs found in the questions although verbs are the most important elements in question designing, e.g. the distinction between 'define the following terms' and 'how would you define the following terms?' is significant here and that depending on a pre-defined list alone might not guarantee accurate performance, particularly for words that are attributed to more than one level (Mohammed and Omar, 2020). The former is associated with the knowledge level while the latter is associated with the comprehension level. Thus, as said earlier, not only syntactic clues were used to indicate the level of the questions, rather semantic clues, and contextual clues were helpful for the categorization as a few questions did not contain keyword verbs and they began with WH-question words such as 'how, what, why, etc.' that were later associated with appropriate levels through meaning and contextual clues. In some cases where the keyword verb was found in the above list, synonymy was helpful, e.g., 'examine' can be synonymous with 'scrutinize, investigate, assess, ...'

3.5 ETHICAL CONSIDERATIONS

Before starting with the data collection, verbal consent was obtained from both deans of the colleges of Nursing and Science, and every individual instructor was contacted via telephone for permission to utilize the questions for the study. After the instructors granted permission, the researchers began collecting the data which lasted more than two months.

4. FINDINGS AND DISCUSSION

The purpose of this section is to present the results obtained from the final exam question papers deduced through descriptive and inferential statistics. It will first analyze the questions into the different cognitive levels of Bloom's Taxonomy and then moves on to the association between the demographic variables and the cognitive levels.

4.1 Findings

Categorization of the Examination Questions into

the Cognitive Levels

What follows illustrates the evaluation of the questions into the cognitive levels of Bloom's Classification by utilizing percentages, as detailed in Table 3 below.

Table 3. Categorization of the Final Examination Questions

Cognitive Levels (BT Objectives)	College of Nursing		College of Science		Total		Level
	F	%	F	%	F	%	F and %
Evaluation	3	1.76	17	4.8	20	3.82	79 (15.08%) High
Synthesis	0	0	11	3.11	11	2.1	
Analysis	16	9.41	32	9.04	48	9.16	
Application	13	7.65	81	22.88	94	17.94	445 (84.92%) Low
Comprehension	41	24.12	57	16.1	98	18.7	
Knowledge	97	57.06	156	44.07	253	48.28	
Total	170	100	354	100	524	100	524

Note. BT= Bloom's Taxonomy

Featured in Table 3 are the results of the question evaluation for both Colleges of Nursing and Science. Accordingly, the knowledge category (f= 253, 48.28%) comprised the highest proportion of all the other question categories. This was followed by the comprehension and application levels, with each containing 98 (18.7%) and 94 (17.94) questions. Additionally, 48 (9.16%) questions belonged to the analysis level. Whereas, merely 20 (3.82%) and 11 (2.1%) questions were attributed to the evaluation and synthesis levels. When considering high-order and low-order thinking questions in both colleges, it becomes clear that the majority of the questions (f=445, 84.92%) dominate the low-order thinking domains, whereas very few questions (f= 79, 15.08%) occur within the domains of high-order thinking

skills.

Further examination of the table reveals striking comparisons between the two colleges in terms of the level of the questions. In both colleges, the knowledge domain comprised the majority of the questions with 97 (57.06%) and 156 (44.07%) questions for the College of Nursing and College of Science respectively. Following this, the comprehension level comprised 41 (24.12%) questions in the College of Nursing while the application level comprised the second highest level in the College of Science with 81 (22.88%) questions.

The Impact of the Demographic Variables on the Level of Questions

To demonstrate the connection between the demographic variables and the quality of the questions, the inferential regression coefficient revealed a significant negative association between the rate of success and the level of the questions.

Table 4. Regression Coefficient for the Association between the Level of the Questions and the Demographic Variables.

Demographic Variables	Un C		St C	t	Sig.
	B	Std. E	Beta		
(Constant)	34.623	7.140		4.849	.000
Rate of success	-.236-	.066	-.407-	-3.565-	.001
Teaching experience	-.042-	.249	-.027-	-.170-	.865
Academic Title	-.960-	2.154	-.070-	-.446-	.657
Academic degree	1.453	2.644	.076	.549	.584

Note. Un C: unstandardized coefficients; St U: standardized coefficients; Std. E: standard error

As the table displays, the only significant correlation was observed between the rate of success and the quality of the questions with the t-value (-3.565) at (0.001). Nevertheless, the correlation of the other demographic variables, i.e., teaching experience, academic title, and academic degree on the level of the questions was not statistically significant.

4.2 Discussion of the Results

One of the characteristics of Bloom's Categorization of cognitive skills involves the devotion of the lower levels of cognitive analysis to objectivity compared to the higher cognitive levels that represent more or less subjectivity although this cannot draw a borderline between high and low levels (Ebadi and Shahbazian, 2015) that might be because the lower level questions are mostly associated with the objective questions including true-false questions, matching questions, multiple-choice questions, and so on, whereas the higher level questions are typically associated with subjective questions such as essay questions whose purpose is to stimulate students to take part in the learning process by personalizing their responses, critical thinking, and self-expression.

The overall results displayed that a large number of the final examination questions revolve around lower-order thinking skills, i.e., knowledge, comprehension, and application, and lacked higher-order cognitive skills suggested by Bloom's Taxonomy. This finding, according to Ebadi and Shahbazian (2015), results from the objective nature of the test items that aim at recalling, understanding, and applying facts. Such tests include close-ended questions or items such as multiple-choice questions, true-false questions, and so on that might have a high degree of practicality and reliability and might produce extrinsic motivation but a less positive washback effect (Ebadi and Shahbazian, 2015). Nevertheless, it has been shown in other studies (Palmer and Devitt, 2007) that, even open-ended questions, if not carefully and properly constructed, could test factual recall similar to multiple-choice questions that can sometimes assess higher-order cognitive skills. Having a limited number of low-order thinking questions is mandatory, particularly for university students because, as explained earlier, Bloom's Taxonomy is hierarchical, meaning that each high level is based on a level below it. However, based on the literature reviewed by Köksal and Ulum, (2018), efficacious exam questions require to encompass sundry cognitive levels to address the

various capacities of learners. Additionally, Naranayan and Adithan (2015) state that, the proportion of high-order thinking questions should record more than 70% of the questions at the university level.

The present study echoes the findings of much prior research in the literature that most test papers contained low-order thinking questions (Baghaei et al., 2020; Ebadi and Shahbazian, 2015; Fayyaz et al., 2019; Kabombwe et al., 2021; Köksal and Ulum, 2018; Palmer and Devitt, 2007; Saido et al., 2015; Swart, 2010). Köksal and Ulum (2018) observed that most of the questions covered knowledge and comprehension levels in almost all the examination components. Congruently, Palmer, and Devitt (2007) showed that even essay questions that are, in essence, employed to assess higher cognitive skills covered only the knowledge level. Fayyaz et al. (2019) further revealed that examiners included low-order thinking skills questions in their examination papers even at the master level. More importantly, Saido et al. (2015) reached the same conclusion that 7th-grade students in Iraqi Kurdistan were at lower levels of thinking skills in a science curriculum and that they require to improve their high-order cognitive skills, especially evaluation and synthesis skills for promoting students' creativity in science. Most importantly, Baghaei et al. (2020) revealed that even the IELTS and TOEFL IBT tests contained more low-order questions on both listening and reading components.

All these congruent results indicate that the inclusion of low-order questions in final examination papers and other contexts seems to be a common, global, and serious issue that requires immediate attention because testing merely factual recall cannot prepare productive students. One legitimate reason behind this is the methods applied in teaching that are essentially more or less traditional methods not stimulating students' reasoning and critical thinking. Another compelling reason for overusing lower-level questions concerns university instructors' explicit awareness of Bloom's Taxonomy which is limited, particularly on the practical side, as supported in the literature that university instructors' awareness of Bloom's Taxonomy is strikingly low (Naranayan and Adithan, 2015).

Another finding reported in this study involved the impact of each of the demographic variables, namely students' rate of success, and instructors' qualification, academic title, and teaching experience

on the level of the questions. The effect of the rate of success was found to be statistically significant, i.e., the rate of success was inversely and negatively associated with the question quality. In other words, the higher the level of the questions, the lower the rate of success, which seems logical because higher cognitive levels require more reasoning, problem-solving, decision-making, and critical thinking skills and can be more challenging. This is supported by Koçdar et al. (2016) that students found the recall and comprehension level questions easier than the application-level questions and obtained higher marks on the low-level questions. Similarly, Swart (2010) obtained the same results that success rates were higher in exam papers that contained low-order questions. As for the instructor-associated variables, i.e., qualification, academic title, and teaching experience, they did not affect question quality, i.e., the level of the questions on Bloom's Taxonomy was the same regardless of whether the instructor had a master's degree or a Ph.D. degree, whether the instructor was an assistant lecturer or a professor, whether he was experienced or inexperienced. Theoretically, this result is unexpected, and instructors with a higher qualification, a higher academic title, and a greater deal of experience should set high-quality questions. The underlying reason behind these insignificant results can be attributed to the fact that newbie teachers are usually more energetic and have more eagerness to include higher-level questions. However, this is not the case in the present study because they do not have enough experience and their qualification and academic title might not help them in designing high-order thinking questions. More importantly, as stated previously, instructors' assessment literacy might be more essential than qualification, academic title, and teaching experience, which is confirmed by Leung (2014). Another reason can be associated with teachers' low awareness of Bloom's Taxonomy, as asserted by Naranayan and Adithan (2015), particularly teachers who are not within the domain of education such as the science specialties. Some of them have knowledge of the taxonomy but they do not know how to apply it. Most importantly, most often the academic titles, the academic degrees, and even the teaching experience do not reflect teachers' academic competence and awareness of the education issues in the context of the present study. Finally, since designing high-order questions requires a lot of effort and is time-consuming for teachers, they might avoid designing these kinds of questions. Above all, the teaching methodologies are more based on the

lower levels; therefore, the exam questions should reflect the teaching methodologies.

5. CONCLUSION AND RECOMMENDATION

The researchers assessed the end-of-semester examination questions at two different colleges based on Bloom's Classification into six cognition levels and then distributed the six levels into two major levels of thinking questions, namely low-order cognitive questions, and high-order cognitive questions. The study concluded that most of the questions concentrated on low-order cognition that lack creativity, reasoning, and critical thinking. Additionally, the relationship between the success rate and the level of the questions was highly statistically significant. The higher the success rate, the lower the question level was, and vice-versa. However, the level of the questions did not vary according to the instructors' teaching experience, qualification, and academic title.

These results presuppose university instructors' low awareness of Bloom's Taxonomy and unbalanced questions. Therefore, this study's recommendations are directed at three different concerned parties involving university instructors, university administrators, and the Ministry of Higher Education and Scientific Research. Firstly, it is recommended that university instructors update their knowledge regarding the more recent methodologies in both instruction and assessment and consider them while teaching and testing. They should prepare tests in a way to identify and indicate what students have learned in the course and check their ability to utilize that knowledge in real-life problems. Secondly, university administrators need to form a committee in each college to inspect the questions for any inaccuracies in both form and content before being printed out and handed to students. Thirdly, the Ministry of Higher Education and Scientific Research in Iraqi Kurdistan, in addition, should encourage universities to focus on the schemes of assessment when designing course syllabuses by including learning objectives that do not only belong to low levels, rather they should belong to higher levels of analysis, synthesis, and evaluation. More importantly, workshops and seminars should be conducted on schemes of assessment, enhancement of high-order thinking skills, and Bloom's Classification to familiarize instructors with the different levels of cognition in both instruction and assessment.

One of the areas that have not been tackled in most educational contexts involves instructors' awareness of Bloom's Taxonomy. Therefore, this study suggests conducting a study regarding instructors' awareness of this categorization of educational objectives. Raising awareness of assessment schemes in general and the various taxonomies of questions can help university instructors design well-balanced and high-quality questions, which is the Cinderella of the current educational research.

6-REFERENCES

- Almerico, G. M. and Baker, R. K., 2004. Bloom's Taxonomy illustrative verbs: Developing a comprehensive list for educator use. *Florida Association of Teacher Educators Journal*, 1(4), pp.1-10. <http://www.fate1.org/journals/2004/almerico1.pdf>.
- Anderson, L.W. and Krathwohl, D., 2001. *A Taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives*. New York: Longman Press.
- Baghaei, S., Bagheri, M. S. and Yamini, M., 2020. Analysis of IELTS and TOEFL reading and listening tests in terms of Revised Bloom's Taxonomy. *Cogent Education*, 7(1), pp.1-23. 10.1080/2331186X.2020.1720939.
- Bloom, B. S., 1956. *Taxonomy of educational objectives*. New York: Longman.
- Chatterjee, S. and Hadi, A. S., 2006. *Regression analysis by example*. John Wiley & Sons.
- Crisp, V., Johnson, M., and Constantinou, F., 2019. A question of quality: Conceptualizations of quality in the context of educational test questions. *Research in Education*, 105(1), pp.18-41. <https://doi.org/10.1177/0034523717752203>.
- Demir, M. K. and Eryaman, M. Y., 2012. A qualitative evaluation instructors' exam questions at a primary education department in terms of certain variables. *Educational Policy Analysis and Strategic Research*, 7(1), pp.52-63. https://dergipark.org.tr/en/pub/epasr/issue/17482/182986#article_cite.
- Ebadi, S. and Shahbazian, F., 2015. Exploring the cognitive level of final exams in Iranian high schools: Focusing on Bloom's Taxonomy. *Journal of Applied Linguistics and Language Research*, 2, 4, pp.1-11. <http://www.jallr.com/index.php/JALLR/article/view/58>.
- Fayyaz, A., Danish, M. H. and UI Hassan, H. H., 2019. Evaluation of MA English questions papers at cognitive level: Application of Bloom Taxonomy. *European Academic Research*, 6(12), pp.7107-7120. <https://euacademic.org/PastIssueList.aspx?artid=74>.
- Himmah, W. I., Nayazik, A. and Setyawan, F., 2019. Revised Bloom's taxonomy to analyze the final mathematics examination problems in junior high school. *Journal of*

- Physics: Conference Series, 8(1), pp.1-7.
<https://doi.org/10.1088/1742-6596/1188/1/012028>.
- Hughes, A., 1989. Testing for language teachers. 2nd ed. Cambridge: Cambridge University Press.
- Jiang, Y., 2020. Teacher classroom questioning practice and assessment literacy: Case studies of four English language teachers in Chinese universities. *Frontiers in Education*, 5, pp.1-17. <https://doi.org/10.3389/educ.2020.00023>.
- Kabombwe, Y., Machila, N. and Sikayomya, P., 2021. A comparative analysis of the Zambian senior secondary history examination between the old and revised curriculum using Bloom's Taxonomy. *Yesterday and Today Journal for History Education in SouthAfricaandAbroad*, 25, pp.1-25. <http://dx.doi.org/10.17159/2223-0386/2021/n25a2>.
- Koçdar, S., Karadağ, N. and Şahin, M. D., 2016. Analysis of the difficulty and discrimination indices of multiple-choice questions according to cognitive levels in an open and distance learning context. *TOJET: The Turkish Online Journal of Educational Technology*, 15(4), pp.16-24.
- Köksal, D. and Ulum, Ö. G., 2018. Language assessment through Bloom's Taxonomy. *Journal of Language and Linguistic Studies*, 14(2), pp.76-88. <http://www.jlls.org/index.php/jlls>.
- Leung, C., 2014. Classroom-based assessment issues for language teacher education. In: A. J. Kunnan, ed. *The companion to language assessment*. Chichester: Wiley Blackwell. pp.1510-1519.
- Abduh, M. Y., 2015. Evaluating examination papers of EFL students at Hodeidah University-Yemen. *Abhath*, 2(3), pp.1-14. <https://ojs.abhathye.com/index.php/OJSABAHA-TH-YE/article/view/321>.
- Mohammed, M. and Omar, N., 2020. Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. *PLoS ONE*, 15(3), pp.1-21. <https://doi.org/10.1371/journal.pone.0230442>.
- Mohamed, O. J., Zakar, N. A. and Alshaikhdeeb, B., 2019. A combination method of syntactic and semantic approaches for classifying examination questions into Bloom's taxonomy cognitive. *Journal of Engineering Science and Technology*, 14(2), pp.935-950.
- Moore, K. D., 2009. *Effective instructional strategies: From theory to practice*. 2 ed. London: Sage.
- Naranayan, S. and Adithan, M., 2015. Analysis of question papers in engineering courses with respect to hots (higher order thinking skills). *American Journal of Engineering Education (AJEE)*. 6(1), pp.1-10.
- Newton, P. M., Da Silva, A. and Peters L. G., 2020. A pragmatic master list of action verbs for Bloom's Taxonomy. *Frontiers in Education*, 5, pp.1-6. [10.3389/educ.2020.00107](https://doi.org/10.3389/educ.2020.00107).
- Omar, N., Haris, S. S., Hassan, R., Arshad, H., Rahmat, M., Zainal, N. F. A. and Zulkifli, R., 2012. Automated analysis of exam questions according to Bloom's taxonomy. *Procedia - Social and Behavioral Sciences*, 59, pp.297 - 303.
- Palmer, E. and Devitt, P. G., 2007. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *Research paper. BMC Medical Education*, 7, 49, 1-7.
- Pappas, E., Pierrakos, O. and Nagel, R., 2012. Using Bloom's Taxonomy to teach sustainability in multiple contexts. *Journal of CleanerProduction*, 48, pp.54-64. <https://doi.org/10.1016/j.jclepro.2012.09.039>
- Richards, J. C. and Schmidt, R., 2010. *Longman dictionary of language teaching and applied linguistics*. 4th ed. New York: Pearson.
- Saido, G. M., Siraj, S., Bin Nordin, A. and Al-Amedy, O. S., 2015. Higher order thinking skills among secondary school students in science learning. *The Malaysian Online Journal of Educational Science*, 3(3), pp.13-20.
- Salkind, N. J., ed., 2010. *Encyclopedia of research design*. London: Sage.
- Shabatura, J., 2013. *Using Bloom's Taxonomy to write effective learning objectives*. University of Arkansas Tips.
- Swart, A. J., 2010. Evaluation of final examination papers in engineering: A case study using Bloom's Taxonomy. *IEEE Transactions on Education*, 53(2), pp.257-264.
- Yahya, A. A., Toukal, Z. and Osman, A., 2012. Bloom's Taxonomy-Based classification for item bank questions using support vector machines. In *Modern Advances in Intelligent Systems and Tools* (pp.135-140). Berlin, Germany: Springer.
- Williams, C., 2007. Research methods. *Journal of Business & Economic Research*, 5(3), pp.125-138.
- Wiseman, D. G. and Hunt, G. H., 2014. *Best practice in motivation and management in the classroom*. 3rd ed. Springfield: Thomas Charles.