Available online at http://aran.garmian.edu.krd

Aran Journal for Languages and Humanities

https://doi.org/10.24271/ARN.2026.02-01- 23

Revisiting Examination Questions within an EFL University Context: Using Bloom's Taxonomy as a Guiding Framework

Karmand Abdulla Hamad

English Department, Faculty of Arts, Soran University, Kurdistan Region - Iraq

Article Info

Received	**2026-12-17**
Accepted	**2026-02-01**
Published:	**2026-02-15**

Keywords

Bloom's taxonomy, examination questions, lower-order thinking skills, higher-order-thinking skills, cognitive progression

Corresponding Author

Abstract:

Despite rising concerns about examinations as a traditional means of assessment, they continue to remain as a dominant tool. This implies that exams are less likely to be abandoned and replaced by other assessment instruments. One major issue with examinations has been that they are excessively obsessed with testing lower-order cognitive skills, especially recalling and comprehending which are identified as surface levels of thinking under Bloom's taxonomy. This does not mean that examinations cannot be applied to measure higher cognitive abilities in learners. On the contrary, exams can well be utilized to assess higher order thinking skills. This, however, largely depends on the quality and level of examination questions. This study, therefore, is an attempt to revisit the cognitive level of examination questions within an EFL university context. The study makes use of Bloom's taxonomy as a framework to guide the analysis and evaluation of examination questions. For this purpose, the final examination questions for two successive academic years 2022-2023 and 2023-2024 were chosen from two English departments within a certain university context. The findings revealed that a significant emphasis of examination questions goes to lower-order cognitive levels, particularly during the early years of the undergraduate study. The emphasis of examination questions, though, gradually shifts toward higher-order thinking levels, especially the analytical skill, at more senior levels of study. The findings also indicated that the focus of examination questions varied across course categories.

Introduction

Assessment practices emerged with the foundation of formal education and establishment of universities which are believed to be several millennia old (Wilbrink, 1997). This implies that assessment has always been an indispensible part of the educational process. Assessment has served different purposes, including gauging student learning, assessing the quality of instruction and informing pedagogical decisions among others (Black & Wiliam, 1998). Multiple assessment tools could be utilized to fulfill these purposes. Among them though, examinations continue to be the most widely used tool to measure the cognitive abilities of students (Sivaraman & Krishna, 2015).

There are multiple reasons that examinations have remained as conventionally and universally dominant forms of assessment. Exams have a long-standing history and are deeply ingrained within educational systems worldwide (Wilbrink, 1997). Exams, particularly standardized tests, are believed to be practically feasible and efficient when a large number of students are intended to be measured (Earl, 2012). At the same time, examinations are believed be more objective and less subjective and biased, especially when consistent grading schemes and rubrics are accounted. That seems to be partly because exams, particularly standardized ones, use the language of numbers and numbers tend to represent objectivity, scientificity and rationality and enable quantification, classification and standardization which are often favored by decision makers and power structures (Shohamy, 2001).

Despite the long-established legacy of exams within various educational contexts, they have been challenged on multiple grounds. One major issue has been overreliance on exams as the sole assessment tool rather than as one single tool among others (Black & Wiliam, 1998). According to French, Dickerson and Mulder (2024), this overreliance on examinations can restrict making use of diverse assessment methods which are necessary for assessing a broad range of learning outcomes, gaining a more holistic picture of students' abilities, encouraging their active engagement and promoting their deeper understandings of the content. Moreover, when exams are applied as the primary assessment tool, they can also diminish opportunities for effective formative assessment and feedback.

Another significant issue of examinations seems to be related to the type and quality of questions they contain. It is argued that examinations still focus disproportionately on lower-order cognitive abilities (Anderson & Krathwohl, 2001). Exam questions that primarily target assessing the abilities of learners to recall and comprehend content are likely to fail to capture the breadth of knowledge, critical thinking and analytical and problem-solving skills among students that education should aim to foster (Anderson & Krathwohl, 2001; Brookhart, 2010). An overemphasis on questions that assess lower-level thinking skills (e.g. remembering and superficial understanding) at the expense of more interpretive, evaluative and critical skills creates a teaching and learning environment where students prioritize memorization and rote-learning (Chandio, Zafar & Solangi, 2021).

Even though lower-order cognitive abilities have their own value and can serve as a basis for the development of higher-order thinking skills, assessment practices should avoid centering entirely on the former (Anderson & Krathwohl, 2001; Gezer, Sunkur & Sahin, 2014). This suggests that a balance needs to be made between questions that intend to measure lower-order thinking skills and those that aim to gauge higher-order thinking skills (Swart, 2009). Educators and researchers often resort to well-established classification systems to address the issue of designing and developing examinations and examination questions. Among them, Bloom's taxonomy seems to be the most widely adopted framework. The taxonomy, developed by Benjamin Bloom and his colleagues in 1956, has an old version with six levels, including Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation (Bloom, 1956).

The taxonomy was later revised by Anderson and Krathwohl in 2001 that modified the terms to become more action-oriented verbs, namely Remember, Understand, Apply, Analyze, Evaluate and Create. This hierarchical framework offers a clear roadmap, particularly for designing instructional practices, setting out educational objectives and learning outcomes and developing examination questions and other assessment practices that can target multiple cognitive levels (Anderson & Krathwohl, 2001; Gezer et al., 2014). Based on the taxonomy and its various classifications, examination questions can either be categorized into lower-order thinking skills (i.e. remembering, understanding and applying) and higher-order cognitive skills (i.e. analyzing, evaluating and creating) or the classification includes three categories, namely lower thinking levels (i.e. remembering and understanding), intermediate thinking levels (i.e. applying and analyzing) and higher thinking levels (i.e. evaluating and creating) (Jones et al., 2009). For the purpose of this study, the latter classification has been adopted with slight modifications, whereby the intermediate thinking skills merely includes the applying level. Thus, the examination questions, under investigation, are analyzed and categorized based on the modified classification.

By adopting Bloom's taxonomy as a guiding framework, the current study seeks to conduct a thorough and systematic analysis of final examination questions within a specific EFL higher education context in Kurdistan Region with an aim of ascertaining the cognitive challenges and demands placed on students. The study also aims to determine any potential imbalances as some teachers may prioritize questions targeting certain cognitive levels of questions over others. Research across various contexts and disciplines (e.g. Freahat & Smadi, 2014; Ebadi & Shahbazian, 2015; Köksal & Ulum, 2018; Fayyaz et al., 2019; Bayaydah & Altweissi, 2020; Chandio et al., 2021; Muhayimana et al.; 2022) have shown that examination questions consistently lean towards assessing lower cognitive abilities. Whereas this trend appears to have been well-documented, continued investigation of other educational environments, especially the ones that have remained underexplored, including the context of this study, seems to be imperative. This study, more specifically, tries to answer the following research questions:

1. Which cognitive levels of Bloom's taxonomy are most widely targeted by examination questions within an EFL university context?
2. How are examination questions distributed across the six cognitive levels of Bloom's taxonomy within the two English departments under study?
3. How do the cognitive demands of examination questions vary across subjects?

Literature Review

Despite the fact that formative and alternative assessment methods are receiving increasing attention, examinations continue to occupy a central position within educational institutions. For that reason, a substantial body of research has tried to examine the quality and cognitive demands of examination questions because the way they are structured and designed reflects not merely assessment practices and beliefs but also underlying pedagogical approaches (Shepard, 2000). This places a demand on teachers to develop their art and skill of questioning so that exams and exam questions can function as effective means of assessment, accommodating distinct learner skills and assessing a broad spectrum of their cognitive abilities (Sivaraman & Krishna, 2015).

The review of the existing literature indicates that the analysis of examination questions has garnered significant scholarly interest across a variety of academic fields and educational contexts. Studies have focused on both school and university levels. Ebadi and Shahbazian (2015) studied the items of English nation-wide final exams of an Iranian high school. The results revealed that of the 55 items of the first grade of high school, 33% of the questions were targeting the knowledge level, 56% were related to the comprehension level and 11% were at the application level. Similar results were found for the second grade questions as 30% of the questions were covering the knowledge level, 57% were at the

comprehension level and 12% aimed to test the application level with no questions assessing analysis, synthesis and evaluation levels. A longitudinal study conducted by Muhayimana, Kwizera, and Nyirahabimana (2022), between 2013-2019 to examine the Rwandan schools Primary Leaving English Exam, found that the questions overwhelmingly belonged to lower-order thinking skills with 98.79% of 574 questions measuring remembering, understanding and applying levels while only 1.21% of the questions focused on higher-order thinking skills.

A similar lengthy research was carried out by Chandio et al. (2021) to analyze the annual question papers designed by the Boards of Intermediate and Secondary Education in Pakistan for the subject of English, grade five from 2014 to 2018. The findings demonstrated that 74% of all questions were devoted to lower-order domains whereas only 26% measured higher-order levels. Bayaydah and Altweissi (2020) studied English final exam questions of 9th and 10th grades in Jordanian secondary schools. Both knowledge and comprehension levels were found to receive the highest attention with 60% while analysis received the least attention with just 4.07%. Another study conducted by Freahat and Smadi (2014), which compared questions within EFL texts from Jordanian universities and secondary schools, yielded a particularly concerning result. The researchers concluded that secondary school textbooks showed a greater emphasis on higher-level thinking questions than university-level textbooks. This is a clear indicator of a lack of cognitive progression and also a possible disconnection between the two educational stages that may leave university students unprepared for future challenging tasks and responsibilities.

University-level examinations and assessments are expected to constantly challenge students with higher-order cognitive tasks and activities. According to Assaly and Smadi (2015), university students should not be simply assessed based on their ability to remember and understand some basic content but rather on their ability to critically examine texts, synthesize ideas and construct logical and well-supported arguments. However, research seems to have shown that as students transition to university education, assessments may not become more cognitively demanding. Köksal and Ulum (2018), in their research of the General English course examination in Turkish universities, revealed that questions entirely focused on testing lower-order cognitive levels among students, lacking questions that target higher level thinking skills. Fayyaz, Danish and Hassan (2019) examined the exam papers of master English courses at Punjab University in Pakistan that included 215 questions of the years 2012 to 2017. The results turned out to be similar to the ones presented by other studies conducted within school contexts and university undergraduate programs with 3.72% of the questions intended to measure knowledge level, 83.72% comprehension level, 10.23% application level and only 1.39% evaluation level.

A more contextually relevant but disciplinarily distant study by Qadir et al. (2023) tried to evaluate the level of examination questions at Colleges of Nursing and Science in Kurdistan Region. The sample of their study included 75 final examination papers that comprised 524 questions. Nearly 85% of the questions seemed to belong to lower-order thinking levels, leaving only 15% focused on higher-order thinking abilities. A study by Virranmäki, Valta-Hulkkonen and Pellikka (2020) investigated the Finnish Entry Examination in Geography based on the new version of Bloom's taxonomy. The number of the questions analyzed included 331 from 2013 to 2019. The results showed that the majority of the questions (over 70%) required conceptual and factual knowledge from students. The results also revealed that following digitalization, questions demanding simple remembering of content have dropped while those that require a higher level of thinking, particularly analyzing, have slightly increased. Letmon (2023) compared and analyzed the cognitive skill levels of high stake examinations of physics and chemistry of six countries, namely England, Ireland, the Netherlands, New South Wales, South Africa and Scotland. The examination questions of 2016 were selected. The analysis demonstrated that across the physics examinations, a greater emphasis was placed on the 'apply' level with a smaller portion of the questions targeting 'analyze' and 'evaluate' levels.

As for the chemistry examinations, a percentage between 83% to 99% was allocated to assessing 'remember', 'understand' and 'apply' levels with very limited number of questions assessing 'evaluate' skill. No exam questions were found to target 'create' level.

The results of the previous studies are a real testament that examinations and assessment practices are overwhelmingly obsessed with lower-order cognitive skills. Research has identified several underlying institutional, curricular and pedagogical factors that have contributed to this continued obsession with surface levels of thinking. One reason could be that teachers are not adequately trained to design higher level questions or that their focus on lower level questions may result from institutional demands and curricular pressures that prioritize content delivery and coverage over an in-depth cognitive engagement of students with the material (Febriyana & Harjanto, 2023; Qadir et al., 2023). Besides, questions that assess simple recall and basic comprehension of content are easier to create and score 'objectively' which makes them rather favored, particularly for large-scale and standardized exams (Wilbrink, 1997; Ebadi & Shahbazian, 2015).

Research Method: Document Analysis

This research adopts document analysis as the primary data collection method. Document analysis, according to Bowen (2009), refers to a methodical tool used for examining and evaluating documents with an aim to elicit meaning, grasp understanding and generate research-based knowledge. Document analysis particularly suits research that seeks to investigate various institutional documents, including policies, reports, curricula, textbooks and examination papers as they allow researchers to uncover what educational philosophies, ideologies, cognitive processes and competencies embedded in written materials (Bowen, 2009; Olson, 2010; Morgan, 2022). As for this study, document analysis began with the selection of relevant documents as the data source for this research which included EFL university examination papers to determine what cognitive demands examination questions largely target.

Materials and Sampling

This study solely relied on document analysis which, according to Bowen (2009), can be used as a 'stand-alone method'. The primary research materials, therefore, consist of existing examination papers chosen from two departments of English, belonging to two different faculties, within a certain higher education institution. The two departments are referred to as Department A and Department B which ensures the anonymity and alignment with ethical research practices. Such kinds of research materials, as Morgan (2022) highlights, are often considered as pre-existing data. Studies that focus on documents for analysis either include "all the relevant documents or takes a sample from a larger number" (Olson, 2012, p. 79). For practical reasons, this study included a certain number of examination papers and questions. The selection of the examination papers was based on purposive sampling which seems to be appropriate for document-based studies (Patton, 2015).

Overall, 120 examination papers were chosen which comprised 945 question items, including both main examination questions and sub-questions. Of this number, 493 question items were from Department A, which comprised 52.17% of the whole dataset, and 452 question items were chosen from Department B which made up 47.83%. There were several criteria upon which examination papers were selected. These include exam types, time frame, study levels and discipline specificity. With respect to exam types, only end-of-semester exam papers were included as they are institutionally more recognized and valued, provide a more comprehensive coverage of content and could be more reflective and representative of pedagogical and assessment beliefs and practices. Regarding the time frame, the study only focused on final examination papers administered during the academic years of 2022-2023 and 2023-2024 as they were more easily accessible and could better capture current practices. The exams were chosen from all the four study levels. This allowed comparing exam questions across different levels based on their

cognitive demands. This also maximized the potential that the results of this study become more representative, generalizable and credible. Finally, the study avoided including exam questions related to general non-English courses that EFL students within Kurdistan higher education contexts are required to take, such as computer science course and kurdology.

Data Analysis Procedure

The process of data analysis for this study followed 'qualitative-quantitative content analysis' which has been referred to as 'mixed content analysis' by certain researchers (Krippendorff, 2004; Marying, 2014). This contains steps of analysis that are both qualitatively and quantitatively driven. Whereas the former is primarily concerned with an in-depth understanding and interpretation of textual data, the latter focuses on numerical analysis of the data. With respect to this study, the qualitative content analysis followed the steps identified by Schreier (2012) and Mayring (2014) that include determining the source of the data (i.e. the examination papers and questions), developing a coding frame or adopting a pre-existing one (i.e. the six thinking levels found in the revised version of Bloom's taxonomy were used as a coding frame in this study), and then treating each examination question as a unit of analysis and assigning each question to one of the six cognitive levels during the main coding phase.

As for the quantitative content analysis, which added an important statistical dimension to the analysis, this involved a systematic quantification of the frequency and distribution of examination questions across the six cognitive levels. This quantitative part of the analysis also allowed to draw numerical comparisons that provided objective insights regarding the cognitive demands embedded in assessment practices. This included the identification of patterns, imbalances or trends across the different courses, levels and departments which revealed the extent to which the examinations try to assess lower, medium or higher levels of thinking.

The analysis of the data revealed examples of questions that were borderline or overlapping. For this reason and also to ensure the reliability and methodological rigor of the study, an experienced researcher (i.e. a second coder) was invited to independently analyze a sample set of questions. Inter-coder reliability, according to Schreier (2012) and Mayring (2014), as a major quality criterion, is of particular significance for content analysis.

Analysis and Results

This section introduces the results obtained from the analysis conducted to the examination questions taken from two departments of English within a university context. The analysis focuses on how the examination questions are distributed across the various cognitive levels of Bloom which was used as an analytical framework. To provide a more systematic and comprehensive account of the results, the analysis was undertaken at three major levels and the results are presented accordingly.

At first, the results, related to the examination questions of each department (i.e. Department A and Department B), are shown separately. This includes examining the cognitive demands of questions within the four study levels of each department as well as their frequency, distribution and comparison which are all guided by Bloom. Second, the analysis turns to demonstrate the results that reveal the similarities and differences that exist at the cognitive level between the examination questions of the two departments. At last, the analysis focuses on the results drawn from the subject classification of the courses. For this purpose, the examination questions from both departments were grouped into four major categories, including language skills courses (e.g. listening and speaking course, reading and writing course, etc.), linguistics-related courses (e.g. morphology, syntax, semantics, etc.), literature-related courses (e.g. modern literature, world literature, literature and ideas, etc.) and pedagogical and language education courses (e.g. diversity education, teaching methods, language assessment, etc.). The question items are,

then, analyzed based on how the questions under each category are distributed across the six cognitive levels of Bloom's taxonomy.

The quantitative content analysis of the examination questions related to Department A, which constituted over 50 percent of all the questions, revealed interesting results about the cognitive levels of examination questions, their distribution, frequency and percentage across the four academic stages. Whereas figures from Table 1 below demonstrates a clear cognitive progression, the data uncovers that lower-order cognitive questions are largely dominant during the course of the four year program.

Table 1. The Distribution and Percentage of Cognitive Levels of Examination Questions in Department A

Levels of Bloom	Year 1		Year 2		Year 3		Year 4	
	Number of Items	pct. %	Number of Items	pct. %	Number of Items	pct. %	Number of Items	pct. %
Remembering	45	47.37%	42	35.59%	29	21.97%	21	14.19%
Understanding	24	25.26%	28	23.73%	33	25.00%	38	25.68%
Applying	19	20%	17	14.41%	21	15.91%	19	12.84%
Analyzing	7	7.37%	20	16.95%	31	23.48%	39	26.35%
Evaluating	0	0.00%	4	3.39%	10	7.58%	21	14.19%
Creating	0	0.00%	7	5.93%	8	6.06%	10	6.76%
Total	95	100%	118	100%	132	100%	148	100%

Data from the above table highlight that first year examination questions predominantly focus on lower-order thinking skills (i.e. Remembering and Understanding) which together constitute the greatest percentage 72.63% of the total number of questions. Applying, which has been considered as an intermediate-order thinking level within the context of this research, seems to have received less emphasis with 20% of the questions. These figures indicate that higher-order thinking skills are marginally present with only 7.37% of the questions, merely targeting Analyzing. No question items were found to represent the other two higher-order thinking levels.

With respect to the second year examination questions, the data reveals that, comparing to the first year questions, the proportion of questions seeking to assess the ability of students to remember content decreases to 35.59%. However, with Understanding which accounts for 23.73%, questions that fall under these two levels still constitute the majority 59.32%. A noticeable change has been observed with regard to Analyzing which increased to 16.95%. When combined, higher-order thinking levels rise to 26.27%.

Examination questions of the third year show further progression towards higher-order cognitive demands which are represented by 37.12% of all the questions. Whereas this can be considered as a positive indicator, as a good number of questions have become analytically more demanding, the fact that lower-order questions are given greater weight of almost 47% of the questions implies that a major shift towards high quality questions seems to be a fundamental challenge for teachers and educational institutions to deal with.

The gradual move of questions towards higher-order cognitive levels turns out to reach the highest point with the fourth year examination questions, accounting for 47.3%. Of the three higher-order thinking skills, Analyzing has the biggest number of questions at 26.35%. On the other hand, lower-order cognitive questions, especially Remembering, markedly decline from 46.32% in the first year to 14.19% in the fourth year while Understanding remains relatively high with 25.68% of the questions.

Whereas this study significantly relied on quantitative evidence to answer the research questions, qualitative evidence has also been integrated to support the quantitative results and also to strengthen

their validity and transparency. This includes a representative sample of examination questions extracted from the actual dataset. The samples represent all the six cognitive levels of Bloom across the four academic years that can be used to verify the accuracy of the analysis and the coding process. Table 2 below contains question samples for all the levels and academic years. Cognitive levels not represented by any questions from the data were labeled as 'None'.

Table 2. Samples of Examination Questions Extracted from the Data of Department A

Levels of Bloom	Academic Year	Sample Questions
Remembering	First	What are the characteristics of short story according to Edgar Allan Poe?
	Second	What is a clause and what are the main types of clauses?
	Third	Define the following syntactic terms.
	Fourth	Define the following terms in Discourse Analysis
Understanding	First	Explain the Five essential parts of plot structure
	Second	explain the literary characteristics of the Modern Literature.
	Third	Explain how adjective phrases can function as complements.
	Fourth	Semantics needs to go beyond language. It is concerned with intra-linguistic relations. Explain this statement in brief.
Applying	First	Write these sentences putting the verbs into the correct tense.
	Second	Change the following ordinary language expressions to poetic ones
	Third	Apply the expansion test to determine if this is Pattern 1: *The sky is blue*
	Fourth	Provide an example for a Nominal Ellipsis in Cohesive Devices.
Analyzing	First	Make a comparison between Tragedy and Comedy
	Second	Elaborate on how setting helps readers understand or analyze a literary text.
	Third	Analyse the following lines from 'She Walks in Beauty' by Lord Byron.
	Fourth	How does New Historicism differ from Old Historicism? Elaborate
Evaluating	First	None
	Second	Discuss the role of women writers in the development of the novel in the Victorian Age.
	Third	Briefly discuss the following literary techniques and phrases reflected in modern literature and explain why they are important.
	Fourth	Pragmatics is the level of language and linguistics that involve human being as the core of its study. Discuss and support your answer with scientific fact.
Creating	First	None
	Second	Write an introductory paragraph to the above topic, express your feelings and emotions.
	Third	Develop a critical essay discussing how the novel reflects the characteristics and themes of the modern novel. Give examples to make this linkage.
	Fourth	Formulate two research questions based on the research title and aims.

Turning now to the results belonging to Department B, the quantitative data emerging from Table 3 below, reveal an even stronger dominance of lower-order cognitive questions, particularly during the initial stages of the program. Similar to the results of Department A, the questions display a systematic shift towards

higher levels of thinking. However, the changes that can be noticed, at the cognitive levels of questions, are evidently insufficient and primarily occur at later stages.

Table 3. The Distribution and Percentage of Cognitive Levels of Examination Questions in Department B

Levels of Bloom	Year 1		Year 2		Year 3		Year 4	
	Question Items	pct. %	Question Items	pct. %	Question Items	pct. %	Question Items	pct. %
Remembering	48	40%	52	43.3%	34	28.3%	14	15.2%
Understanding	34	28.3%	40	33.3%	39	32.5%	21	22.8%
Applying	30	25%	22	18.3%	22	18.3%	22	23.9%
Analyzing	8	6.7%	6	5.1%	16	13.3%	17	18.5%
Evaluating	0	0.0%	0	0.0%	7	5.8%	16	17.4%
Creating	0	0.0%	0	0.0%	2	1.8%	2	2.1%
Total	120	100%	120	100%	120	100%	92	100%

Data, pertinent to year one of the study, clearly show that lower-order cognitive skills are given greater priority over other skills with 68.3% of all the questions. The level of thinking with the most stable percentages throughout the four academic years turned out to be the Applying level which has 25% of the questions of first year. The other three levels are minimally present through Analyzing, accounting for only 6.7%. The way questions are distributed at this stage provides ample evidence that remembering and comprehending factual content are strongly emphasized.

Contrary to expectations, questions of year two, to a greater extent, lean towards lower-order cognitive levels. The percentage of questions devoted to testing students' ability to remember and understand move upwards to 76.6%. This has led to the decline of questions that require students to apply the knowledge they learn to 18.3% (the same figure observed with the third-year questions). As a result, merely 5.1% of the questions at this stage are categorized as higher-order cognitive questions, entirely allocated to the Analyzing level. That leaves no questions to represent the Evaluating and Creating levels.

With regard to the questions of year three, they mark a more noticeable transition towards higher-order cognitive expectations with 20.9% of questions. Among the three higher cognitive levels, Analyzing has the highest share of questions at 13.3%. This percentage, however, still looks relatively low compared to questions that are classified as lower-order levels which constitute almost two-thirds of the total number of questions at 60.8%, divided between Remembering 28.3% and Understanding 32.5%.

The cognitive levels of the fourth-year examination questions appear to be substantially different from previous years. Questions targeting Remembering and Understanding levels become less prominent with 15.2% for the former and 22.8% for the latter. Applying moderately rises to 23.9% which can be considered as an indication that the practical use of knowledge by students remain as a priority, especially within the context of Department B. Importantly, more cognitively demanding questions constitute a significantly higher proportion with 38% which is equal to the number of questions categorized as lower-order levels. Questions falling under the Analyzing level, which has been more or less present across the four study levels, go up to 18.5%. Meanwhile, a dramatic increase can be seen with the questions aiming at assessing the evaluating ability of learners, reaching 17.4%. Lastly, questions identified at the Creating level remain at a low level with only 2.1%.

Similar to Department A, illustrative examples of examination questions, related to Department B, are also provided below.

Table 4. Samples of Examination Questions Extracted from the Data of Department B

Levels of Bloom	Academic Year	Sample Questions
Remembering	First	Poetry can be defined as …… .
	Second	Define the terms "Subculture', "Dictatorship", "Acculturation".
	Third	What is Connectivism theory?
	Fourth	Mention at least two factors that can affect the motivation of EFL learners.
Understanding	First	Summarize the third paragraph above.
	Second	Social learning is a theory of learning, explain how learning occurs according to this theory?
	Third	What are the two distinctive features of derivational morphemes? Illustrate.
	Fourth	What is your understanding of "diverse learning needs?
Applying	First	Change the following sentences from Active to Passive
	Second	Write the transcription of the following words.
	Third	Write the part of speech for following words.
	Fourth	Identify the head noun, pre-modifiers, post-modifiers in the following sentences.
Analyzing	First	How do you make a distinction between fact and opinion in a text? Explain it, and support your answer with providing examples
	Second	Identify two differences between Prose and Poetry.
	Third	Is cultural transmission a feature of animal communication? If not, how do animals learn to communicate?
	Fourth	Analyze the following text semantically (types of meaning, lexical relations, semantic features)
Evaluating	First	None
	Second	None
	Third	Express your view on the above statement by Safa Xlusy on translation.
	Fourth	Embracing diversity is a cornerstone of modern education. How do you think embracing and respecting diversity in the classroom may affect the overall achievement of your students?
Creating	First	None
	Second	None
	Third	Translate the following text into English.
	Fourth	Provide a full sentence example for each function below.

Now that the results related to both departments are shown above, it is time to present how the cognitive demands of examination questions have changed and progressed within the two departments across the four academic years. Table 5 below displays a comparative distribution of the examination questions based on the six cognitive levels of Bloom. The comparison illustrates the degree of change at each level of thinking across the four years and also between the departments.

Table 5. Comparative Analysis of the Distribution of Examination Questions in Department A and Department B

Levels of Bloom	Department	Year 1	Year 2	Year 3	Year 4
Remembering	Department A	47.37%	35.59%	21.97%	14.19
	Department B	40%	43.3%	28.3%	15.2%

Understanding	Department A	25.26%	23.73%	25%	25.68%
	Department B	28.3%	33.3%	32.5%	22.8%
Applying	Department A	20%	14.41%	15.91%	12.84%
	Department B	25%	18.3%	18.3%	23.9%
Analyzing	Department A	7.37%	16.95%	23.48%	26.35%
	Department B	6.7%	5.1%	13.3%	18.5%
Evaluating	Department A	0.0%	3.39%	7.58%	14.19%
	Department B	0.0%	0.0%	5.8%	17.4%
Creating	Department A	0.0%	5.93%	6.06%	6.76%
	Department B	0.0%	0.0%	1.8%	2.1%

A quick look at the table reveals a gradual, but perhaps insufficient and at times inconsistent, decline in the number of questions that are aligned with lower-order cognitive levels. As for the higher-order thinking skills, a moderate progress can be observed towards questions that require deep cognitive engagement. That appears to be particularly true with the Analyzing and Evaluating levels.

Another important observation that can be drawn from the table is that the transition from lower-order to higher-order cognitive levels seems to take place earlier and at a greater extent with Department A than with Department B, where the changes come relatively at later stages with a lesser extent. The following table presents further comparative results drawn from the data combined across the four years of study for both departments.

Table 6: Overall Comparisons of Cognitive Levels (All the Four Years Combined)

Levels of Bloom	Department A	Department B
Remembering	27.59%	27.71%
Understanding	24.95%	29.00%
Applying	15.21%	19.70%
Analyzing	19.68%	14.72%
Evaluating	7.51%	7.14%
Creating	5.07%	1.73%

Although the data put forward earlier about each department separately and about the different levels of study showed notable discrepancies, the overall distribution of questions across the various levels of Bloom's taxonomy indicates that examination questions are predominantly inclined towards testing the low levels of thinking among students. That turned out to be the case for both departments. As the numbers suggest, questions, categorized under Remembering and Understanding combined, constitute more than half of the entire dataset.

Contrasting results are slightly more apparent between the two departments at the intermediate and higher thinking levels. At the intermediate level, Department B places greater emphasis on Applying with nearly 20% of questions throughout the four years. This cognitive level seems to receive less attention by Department A. Whereas, at the higher level, the two departments put almost equal weight on Evaluating at around 7%, examination questions targeting the Analyzing and Creating levels are more widely present within Department A. This disparity can be partly attributed to the nature of the courses studied within the two departments which is further examined below.

Table 7 below shows the distribution of question items and their percentages based on the four course categories. Question items related to literary subjects have the highest number with 301 of questions,

followed by linguistic subjects with 262 questions. Questions identified under language skills courses come third with 238 questions. The least number of questions goes to pedagogical and language education courses with 144 question items.

Table 7. The Distribution of Question Items based on Course Categories

Course Categories	Question Items	Percentage
Language Skills Courses	238	25.19%
Linguistics-related Courses	262	27.72
Literature-related Courses	301	31.85%
Pedagogical and Language Education Courses	144	15.24%
Total	945	100%

Table 8 reveals interesting results of how the question items of each course category are distributed across the six cognitive levels of Bloom's taxonomy.

Table 8: The Distribution of Question Items belonging to the Course Categories across the Six Cognitive Levels of Bloom's Taxonomy

Levels of Bloom	Language Skills Courses		Linguistics-related Courses		Literature-related Courses		Pedagogical & Language Education Courses	
	Question Items	pct. %	Question Items	pct. %	Question Items	pct. %	Question Items	pct. %
Remembering	82	34.45%	63	24.05%	64	21.26%	24	16.67%
Understanding	72	29.83%	66	25.19%	71	23.59%	38	26.39%
Applying	69	28.99%	58	22.14%	31	10.30%	41	28.47%
Analyzing	12	5.04%	52	19.85%	77	25.58%	23	15.97%
Evaluating	4	1.68%	15	5.73%	40	13.29%	14	9.72%
Creating	0	0.0%	8	3.05%	18	5.98%	4	2.78%
Total	238	100%	262	100%	301	100%	144	100%

The results presented earlier regarding the two departments revealed that lower-order cognitive questions are significantly favored over higher-order cognitive questions throughout the dataset. A similar pattern can be seen with this round of analysis. The difference that is observed from the table can be attributed to the nature and type of the courses from which the examination questions were drawn.

One striking result emerging from the table shows apparent disciplinary differences with the cognitive levels of examination questions across the four course categories. Courses, with their focus on language skills, are primarily oriented towards lower-order cognitive questions with extremely marginal representation of higher level cognitive questions. Courses, classified under pedagogy and language education, seem to occupy an intermediate position. The biggest proportion of questions are meant to target the Applying level which indicates a great emphasis on the application of pedagogical knowledge within real-world contexts. At the same time, a moderate increase of questions targeting higher-level thinking can be noticed. Despite the fact that lower level questions continue to dominate linguistics-related courses, a considerable increase can be observed in higher-order thinking questions, especially at the Analyzing level which makes the other two higher-order cognitive levels (i.e. Evaluating and Creating) comparatively low. The most striking shift of examination questions towards higher cognitive levels appears to happen with literary courses where questions requiring advanced cognitive engagement

outnumbered those assessing simple recall and basic comprehension. Of the three higher-order thinking levels, question items classified under the Analyzing and Evaluating levels dramatically increased, demonstrating the priority given to assessing analytical, interpretive, evaluative and critical skills among students.

Discussion

This study was an attempt to investigate the cognitive levels of examination questions within an EFL university context through the lens of Bloom's taxonomy which was adopted as a guiding framework. By aligning the results of the current study with the ones put forward by previous studies, this section intends to situate this research within wider discussions around assessment practices, cognitive skills and development and pedagogical and epistemological alignments.

One outstanding finding regarding the intellectual demands of the existing EFL examination questions was the predominance of lower-order cognitive questions which seems to closely align with research conducted within various educational levels and contexts (e.g. Freahat & Smadi, 2014; Ebadi & Shahbazian, 2015; Köksal & Ulum, 2018; Fayyaz et al., 2019; Bayaydah & Altweissi, 2020; Chandio et al., 2021; Muhayimana et al., 2022). This finding directly addresses the first research question that EFL examination questions disproportionately target lower-order thinking skills, especially remembering and understanding. Data from the two departments and across the distinct study levels altogether revealed that these two cognitive levels of Bloom comprise more than half of the whole dataset.

This persistent dominance of lower cognitive questions, which has also raised widespread concern, seems to reflect certain deeply-rooted perspectives of knowledge, teaching and assessment. According to Knight, Shum and Littleton (2014), these elements are in a 'triadic relationship' with one another and that epistemological assumptions fundamentally shape pedagogical and assessment practices. The fact that the majority of examination questions intended to assess lower thinking abilities of students provides insights on the underlying epistemological perspectives that give priority to certain pedagogical methods and forms of assessment over others.

When assessment tools and among them examinations primarily seek to objectively measure the amount of content or knowledge students acquire, the prevailing epistemological view evident here is associated with positivism (Elkind, 1997; Nieminen & Lahdenperä, 2024). This paradigmatic position views knowledge as an object or commodity that can be transferred from the teacher to the learner (Short & Burke, 1994). This characterizes the teacher as the carrier and dispenser of knowledge while positions the learner as the passive recipient of that knowledge. As a result, assessment mainly circulates around measuring how well learners are capable of reproducing or providing some basic explanation about the content delivered to them (Serafini, 2000). This overemphasis placed on surface-level thinking uncovers the influence that the positivistic understanding of knowledge has on assessment and examination practices while at the same time profoundly shaping pedagogical beliefs and approaches which seem to favor a 'transmissive' model of teaching.

Another important finding that stood out from the results which is apparently relevant to the second research question is that a slow but steady progression towards higher levels of cognition was observed in examination questions within both departments and across the academic years. This was particularly true for Department A as the shift was comparatively more evident and to a greater extent than Department B where the progression emerged later and to a lesser extent. On the one hand, this finding differs from several previous studies (e.g. Freahat & Smadi, 2014 Köksal & Ulum, 2018; Fayyaz et al., 2019) which reported no significant evidence of higher cognitive questions at different levels of study. On the other hand, it seems to be consistent with what Assaly and Smadi (2015) assert that examination questions at

university level should continually move towards assessing deep and cognitively challenging levels of analyzing, evaluating and creating.

The gradual increase of higher-order thinking questions from year one to year four appears to partially align with expectations. However, the proportion of higher level questions remained insufficient and unbalanced as lower cognitive questions were still substantially present even at senior levels of study. As Anderson and Krathwohl (2001) and Gezer et al. (2014) highlight, lower-order thinking skills, particularly understanding, can work as a foundation upon which higher-order cognitive skills can be developed. A serious problem, though, can arise when these low thinking skills become the ultimate goal of the assessment and examination processes which turned out to be relatively the case within the context under investigation. This obviously contradicts the principle of hierarchy which is fundamentally important to Bloom's taxonomy (Anderson & Krathwohl, 2001). Following this principle, examination questions are expected to be hierarchically structured to include low, intermediate and high cognitive questions, with their distribution principally dependent on the level of study. Therefore, when examination questions largely lean towards measuring surface levels of thinking throughout the years of study, then transition from one academic level to another may not necessarily lead to substantial cognitive development.

On the question of how the cognitive demands of examination questions vary across subjects, this study found that examination questions and their cognitive demands are strongly connected to the epistemic nature of disciplines. Examination questions drawn from courses related to language skills were found to be overwhelmingly tilted towards remembering and understanding which together with applying constituted almost all examination questions, leaving roughly limited or no room for higher-order cognitive questions. This finding resonates with what Ebadi and Shahbazian (2015) and Köksal and Ulum (2018) found that language skills courses heavily emphasized recalling and comprehending factual material. One reason that examination questions under this course category almost entirely fall under lower thinking skills could be that such courses seem to predominantly focus on language forms and accuracy which sounds unjustifiable given the fact that contemporary language teaching approaches pay enormous attention to skillful, authentic and reflective language use.

Perhaps the most striking finding, particularly with regard to the third research question, was that examination questions associated with literary courses were outstandingly oriented towards deeper levels of thinking. This supports the argument put forward by other researchers (e.g. Lazere, 1994; Assaly & Smadi, 2015) that literature and literary subjects are naturally more conducive to the integration of higher-order skills. Under this academic discipline and perhaps for the first time across the dataset, higher-order cognitive questions exceeded lower-order cognitive questions where students were expected to be more deeply engaged with the analysis and evaluation of texts and materials under study. This, however, was not the case with all literary courses. On the contrary and as the results displayed, a big proportion of question items, classified under this course category, was meant to test students' lower levels of thinking.

The above discussion explicates that while the nature of courses matters, as some more readily lend themselves to higher cognitive skills than others, elevating the cognitive demand of examination questions hugely rely on teachers, their pedagogical and assessment beliefs and practices as well as their familiarity with models like the Bloom's taxonomy. This enables them to design and structure their assessment and questions more cautiously and effectively that can help their students develop higher-level thinking skills. Therefore, regardless of the courses they teach, teachers should imperatively try to, at least, strike a balance between lower and higher thinking skills, and where feasible, put a greater weight on the latter. However, as Wilbrink (1997) points out, recall-based questions are easier to create and mark which makes them more tempting. Hence, teachers are less likely to take major steps towards incorporating more higher-order and fewer lower-order cognitive questions.

For the assessment culture to go through a vital change, steps are needed to be taken at pedagogical and epistemological levels or more specifically with respect to the key elements of constructive alignment (Biggs, 2003). This implies that developing the quality of assessment, including examination questions, depends on designing learning outcomes that aim to foster higher cognitive abilities and adopting compatible pedagogical methods and strategies that empower students to achieve cognitively demanding learning outcomes. The fact that the findings of this study uncovered some serious gaps with the quality and level of EFL examination questions within a university context can also be considered as a clear sign that there are systemic issues with the other major components of constructive alignment that dynamically shape one another. The teacher, who should be helped to be prepared pedagogically, academically and professionally for this, can play a pivotal role to constructively align these major ingredients with one another that subsequently contributes to the learning process and to the intellectual growth of learners.

Conclusion

This study set out to examine the level of examination questions based on Bloom's taxonomy. The study corroborates the conclusion, consistently drawn across the existing literature, that examination questions predominately centered on lower-order cognitive skills while higher-order cognitive skills, either intentionally or unintentionally, have been overlooked to a worrying extent. What exacerbates the concern is that, contrary to expectations, this unequal distribution of questions, whereby surface-levels of thinking receive greater weight, even happen within higher education institutions where students are supposedly required to engage in deep, analytical and critical thinking and questioning. Nevertheless, the new insights emerged from the study regarding the gradual cognitive progression at disciplinary and departmental levels towards higher thinking skills should help us realize that, although attempts are limited, they deserve recognition, encouragement and support. The findings suggest that for examinations and assessment practices to reflect university education, urgent actions need to be taken by different stakeholders. At institutional level, universities should initiate with more targeted teacher education programs to help teachers enhance their assessment and pedagogical literacy. At a personal level, teachers are required to step forward to try out more innovative pedagogical and assessment tools so that they can align their teaching and assessment practices with the current needs and developments and to fulfill the expectations of their students.

Reference List

1. Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc..
2. Assaly, I. R., & Smadi, O. M. (2015). Using Bloom's Taxonomy to Evaluate the Cognitive Levels of Master Class Textbook's Questions. *English Language Teaching*, *8*(5), 100-110.
3. Bayaydah, A. M., & Altweissi, A. I. (2020). A Bloom's Taxonomy-based Analysis of 9th and 10th Grades English Language Textbooks' Final Examinations and Revision Questions. *International Online Journal of Primary Education*, *9*(2), 197-211.
4. Biggs, J. (2003). Constructing learning by aligning teaching: Constructive alignment. *Teaching for quality learning at university*, *1*(4), 11-33.
5. Black, P., & Wiliam, D. (1998). *Inside the Black Box: Raising Standards Through Classroom Assessment*. Phi Delta Kappan, 80(2), 139-148.
6. Bloom, B. S. (1956). *Taxonomy of Educational Objectives*, *250*.
7. Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, *9*(2), 27-40.
8. Brookhart, S. M. (2010). Assessing higher order thinking skills: An overview. *Practical Assessment, Research & Evaluation*, 15(7).
9. Chandio, M. T., Zafar, N., & Solangi, G. M. (2021). Bloom's Taxonomy: Reforming Pedagogy through Assessment. *Journal of Education and Educational Development*, *8*(1), 109-140.
10. Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin press.
11. Ebadi, S., & 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), 1-11.
12. Elkind, D. (1997). The death of child nature: Education in the postmodern world. *The Phi Delta Kappan*, *79*(3), 241-245.
13. Fayyaz, A. M. N. A., Danish, M. H., & Hassan, H. H. (2019). Evaluation of MA English question papers at cognitive level: Application of Bloom Taxonomy. *European Academic Research*, *6*(12), 7107-7120.
14. Febriyana, F., & Harjanto, I. (2023). Cognitive levels of questions by Indonesian teachers of English. *Journal of English Language Teaching and Linguistics*, *8*(2), 2023.
15. Freahat, N. M., & Smadi, O. M. (2014). Lower-order and higher-order reading questions in secondary and university level EFL textbooks in Jordan. *Theory and Practice in Language Studies*, *4*(9), 1804-1813.

16. French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, *88*(3), 893-918.
17. Gezer, M., Sunkur, O. M., & Sahin, I. F. (2014). An Evaluation of the Exam Questions of Social Studies Course According to Revised Bloom's Taxonomy. *Education Sciences & Psychology*, *28*(2).
18. Jones, K. O., Harland, J., Reid, J. M., & Bartlett, R. (2009). Relationship between examination questions and Bloom's Taxonomy. In *2009 39th IEEE frontiers in education conference* (pp. 1-6). IEEE.
19. Knight, S., Shum, S. B., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. *Journal of Learning Analytics*, *1*(2), 23-47.
20. Köksal, D., & Ulum, Ö. G. (2018). Language assessment through Bloom's Taxonomy. *Journal of language and linguistic studies*, *14*(2), 76-88.
21. Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, *30*(3), 411-433.
22. Lazere, D. (1994). Critical thinking in college English studies. *Inquiry: Critical Thinking Across the Disciplines*, *14*(1), 84-88.
23. Letmon, D. (2023). *Examining the cognitive demand of high-stake physics and chemistry examinations in Ireland* (Doctoral dissertation, Dublin City University).
24. Marying, P. (2014). Qualitative Content Analysis-Theoretical Foundation. *Basic Procedures and Software Solution*, 154-169.
25. Morgan, H. (2022). Conducting a qualitative document analysis. *The qualitative report*, *27*(1), 64-77.
26. Muhayimana, T., Kwizera, L., & Nyirahabimana, M. R. (2022). Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools. *Curriculum Perspectives*, *42*(1), 51-63.
27. Nieminen, J. H., & Lahdenperä, J. (2024). Assessment and epistemic (in) justice: how assessment produces knowledge and knowers. *Teaching in Higher Education*, *29*(1), 300-317.
28. Olson, M. (2010). Document analysis. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research* (pp. 318–320). Sage Publications.
29. Olson, W. K. (2012). Document analysis. In *Data collection: Key debates and methods in social research* (pp. 79-82). Sage Publications Ltd.
30. Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.
31. Qadir, S. M., Omar, R. M., Rasheed, M. H., & Mohammed, C. J. (2023). Assessing the end-of-semester examination papers during the implementation of the bologna process: bloom's taxonomy as a framework. *Koya University Journal of Humanities and Social Sciences*, *6*(1), 77-87.
32. Schreier, M. (2012). Qualitative content analysis in practice. Sage Publications.
33. Serafini, F. (2000). Three paradigms of assessment: Measurement, procedure, and inquiry. *The Reading Teacher*, *54*(4), 384-393.
34. Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*, 29(7), 4-14.
35. Shohamy, E. (2001). Democratic assessment as an alternative. *Language testing*, *18*(4), 373-391.
36. Short, K., & Burke, C. (1994). *Creating curriculum*. Portsmouth, NH: Heinemann.
37. Sivaraman, S. I., & Krishna, D. (2015). Blooms taxonomy–application in exam papers assessment. *Chemical Engineering (VITU)*, *12*(12), 5-9.
38. Swart, A. J. (2009). Evaluation of final examination papers in engineering: A case study using Bloom's Taxonomy. *IEEE Transactions on Education*, *53*(2), 257-264.
39. Virranmäki, E., Valta-Hulkkonen, K., & Pellikka, A. (2020). Geography tests in the Finnish Matriculation Examination in paper and digital forms–An analysis of questions based on revised Bloom's taxonomy. *Studies in Educational Evaluation*, *66*, 100896.
40. Wilbrink, B. (1997). Assessment in historical perspective. *Studies in Educational Evaluation*, *23*(1), 31-48.