



University of
Sistan and Baluchestan



Uppsala University

Cognitive Trends in Iranian High-School Final Examinations: A Bloom's Taxonomy-Based Longitudinal Analysis of Persian Language and Literary Studies and Techniques (2019–2024)

Ebrahim Samani, Ghasem Mehravar Giglou, Abdolhossein Heydari

Assistant Professor, Department of English Language Teaching, Farhangian University, Tehran, Iran, Email: e.samani@cfu.ac.ir

Assistant Professor, Department of Persian Language and Literature Education, Farhangian University, Tehran, Iran, Email: gh.mehravar@cfu.ac.ir

Assistant Professor, Department of English Language Teaching, Farhangian University, Tehran, Iran, Email: a.heidari@cfu.ac.ir

Abstract

This study examines the cognitive characteristics of Iranian high-school final examinations over the six years (2019–2024) using Bloom's taxonomy of cognitive processes. The design of the study was a quantitative content analysis with qualitative illustrative examples. All Persian Language and Literary Studies and Techniques questions were coded according to the revised Bloom's six cognitive levels—Remember, Understand, Apply, Analyze, Evaluate, and Create. Descriptive statistics measured yearly distributions, and Chi-Square Test tested associations between examination year and cognitive-level patterns. Results revealed a dominant focus on lower-order cognition: approximately 80 % of questions targeted the Remember, Understand, or Apply levels, while Evaluate and Create were almost absent. Chi-Square Test did not confirm a significant relationship between year and higher-order cognitive-level distribution. The sustained dominance of rote-learning-oriented items indicates limited emphasis on critical thinking or creative problem solving in Iran's assessment system. The study concludes that systematic reform in examination design and teacher training is needed to promote analytical and creative engagement.

Keywords: Bloom's taxonomy, cognitive levels, final examinations, question analysis, chi square test, critical thinking, Iranian education

I. Introduction

Examinations play a pivotal role in shaping the quality, direction, and perception of educational systems. In many countries, including Iran, national final examinations determine not only students' academic achievements but also their access to higher education and future opportunities. As a result, the cognitive depth and intellectual demand of exam questions serve as important indicators of the effectiveness of the educational process. The degree to which assessments move beyond rote recall toward reasoning, analysis, and creativity reflects how well a curriculum cultivates the higher-order thinking skills necessary for participation in modern, knowledge-based societies.

Within this context, Bloom's taxonomy of cognitive objectives has remained one of the most influential frameworks for evaluating the intellectual rigor of educational assessment. First introduced by Benjamin Bloom and colleagues in 1956 and later revised by Anderson and Krathwohl in 2001, the taxonomy provides a hierarchical model for categorizing cognitive processes involved in learning: remembering, understanding, applying, analyzing, evaluating, and creating. These levels represent a continuum from lower-order thinking (LOT) to higher-order thinking (HOT). Because examinations largely determine what teachers emphasize and what students prioritize, understanding how exam questions are distributed across these levels can reveal much about the implicit pedagogical orientation of a national education system.

In Iran, final examinations at the high-school level—particularly those administered for Grade 12—carry high stakes. They are standardized, centrally designed by the Ministry of Education, and directly influence university admissions. Consequently, the structure and quality of these exams significantly affect classroom teaching and learning practices. Despite curricular reforms intended to promote creativity, reasoning, and critical thinking, anecdotal and empirical evidence suggests that the examinations may still emphasize memorization and surface-level comprehension. Such imbalance between curriculum intentions and assessment practices raises questions about whether the assessment system truly supports the development of higher cognitive competencies.

Over the past decade, educational researchers in Iran have conducted several isolated studies analyzing exam questions using Bloom's taxonomy. Most of these studies have focused on a single subject or a limited time frame, reporting a consistent tendency toward lower cognitive levels. For example, analyses of science, mathematics, and language tests frequently show that the majority of items assess remembering or understanding rather than analysis or evaluation. However, these studies have rarely been longitudinal, and few have compared multiple core subjects over an extended period. Consequently, the national trend in the cognitive demand of exam questions across years and subjects remains largely unexplored.

To address this gap, the present study investigates the cognitive distribution of questions from two major subjects in Iran's Grade-12 national examinations—Persian Language and Literary

Studies and Techniques —over a six year period (2019–2024). These two subjects were selected because together they constitute the core of linguistic and literary education in Iranian secondary schools. While the Persian Language course focuses on grammar, comprehension, and writing, Literary Studies and Techniques course emphasizes literary analysis, appreciation, and aesthetics. Examining both simultaneously provides a holistic view of how Iran’s educational assessment system operationalizes cognitive objectives in the humanities domain. By merging analyses of Persian Language and Literary Studies and Techniques questions, it seeks to reveal whether a meaningful trend has emerged over time toward deeper, more complex forms of cognition. In doing so, it adds to the global conversation on how educational systems can balance factual knowledge with analytical reasoning, ensuring that assessment practices genuinely support the cultivation of critical and creative learners. The present investigation seeks to answer two main research questions:

1. What is the distribution of cognitive levels, based on Bloom’s taxonomy, in the final examination questions of Persian Language and Literary Studies and Techniques courses from 2019 to 2024?
2. Is there a statistically significant relationship between examination year and presence and absence of higher-order thinking, indicating a trend toward higher-order thinking over time?

2. Review of the Literature

2.1. The Centrality of Examination Analysis in Educational Quality

Educational assessment is widely recognized as a mirror of curriculum effectiveness and instructional priorities. According to Haladyna (2004), test items are the most concrete indicators of what teachers value and what students are expected to master. In national examination systems, particularly those tied to high-stakes decision-making, the nature of exam questions strongly influences classroom practices, shaping both teaching strategies and student learning behaviors (Gulikers, Kester, Kirschner, & Bastiaens, 2008). When assessments primarily require factual recall, they tend to encourage surface learning; when they demand interpretation, synthesis, or evaluation, they foster deeper learning and critical reasoning (Biggs & Tang, 2011).

In many countries, centralized examinations are used to maintain fairness, comparability, and accountability across schools. Yet such examinations often prioritize reliability and scoring efficiency over the measurement of complex cognitive skills (Stiggins, 2017). The challenge, therefore, lies in balancing the standardization required for large-scale assessment with the need to evaluate students’ higher-order cognitive abilities. The Iranian national final examinations, which determine secondary-school graduation and university eligibility, epitomize this tension. Despite curricular reforms advocating creativity and problem solving, research has repeatedly shown that these exams remain heavily weighted toward low-level cognitive tasks.

2.2. Bloom's Taxonomy as a Framework for Analyzing Cognitive Demand

Bloom's Taxonomy of Educational Objectives (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) remains the most enduring framework for categorizing the cognitive processes involved in learning and assessment. The taxonomy conceptualizes cognition as hierarchical, ranging from lower-order levels—Knowledge, Comprehension, and Application—to higher-order levels—Analysis, Synthesis, and Evaluation. Each level represents a qualitatively distinct kind of mental operation. While the original taxonomy emphasized cumulative mastery, its revised version by Anderson and Krathwohl (2001) restructured the hierarchy into two dimensions: the knowledge dimension (factual, conceptual, procedural, and metacognitive) and the cognitive process dimension (remember, understand, apply, analyze, evaluate, and create). This revision better captures the interactive and dynamic nature of cognition and learning.

Using Bloom's taxonomy to analyze exam questions provides insight into the cognitive load demanded by assessment tasks. Numerous studies (e.g., Krathwohl, 2002; Forehand, 2010) have emphasized that a balanced distribution across all levels is ideal, ensuring that assessments capture not only factual understanding but also reasoning, problem solving, and creativity. The taxonomy has thus become a standard tool for evaluating the alignment between curriculum, instruction, and assessment—a concept known as constructive alignment (Biggs, 2003). When exams fail to reach higher levels, constructive alignment breaks down, leading to a disconnect between intended learning outcomes and what is actually measured.

A major conceptual contribution of Bloom's taxonomy is its alignment with modern theories of critical thinking. Facione (1990) identified six core skills—interpretation, analysis, evaluation, inference, explanation, and self-regulation—that correspond closely to Bloom's higher cognitive levels. Lipman (1988) argued that evaluation and synthesis are at the heart of critical-thinking pedagogy. Consequently, exam systems emphasizing only recall and comprehension inevitably discourage the very reasoning processes that contemporary education seeks to promote.

Within Iran's evolving educational landscape, this linkage is especially salient. The country's Vision 2025 strategic plan and subsequent curricular reforms explicitly highlight creativity, innovation, and reasoning as national educational goals (Supreme Council of Education, 2012). Yet the effective realization of these objectives depends on whether assessment tools—particularly national examinations—reflect them. As noted by Ebadi and Shahbazian (2015), when assessments remain dominated by low-order cognitive items, teachers are unlikely to adopt methods that cultivate analytical or evaluative thinking. Thus, aligning exam content with Bloom's upper levels is essential not only for accurate evaluation but also for pedagogical transformation.

2.3. International Studies on Exam Cognition and Bloom's Taxonomy

Since the early 2000s, a substantial body of research has examined school and university examinations using Bloom's taxonomy and related frameworks of cognitive demand. These studies have frequently reported an imbalance between lower- and higher-order cognitive processes in formal examinations. For example, Muhayimana, Kwizera, and Nyirahabimana (2022) analyzed 574 questions from Rwanda's Primary Leaving English Examinations using the revised Bloom's taxonomy and found a strong predominance of lower-order thinking skills, which accounted for 98.79% of the items, while only 1.21% required higher-order thinking. Their findings illustrate how high-stakes examination systems may remain heavily concentrated on recall, comprehension, and other lower-level cognitive processes despite curricular aspirations to develop more complex thinking.

Comparable patterns have been reported in other educational contexts. Chowdhury et al. (2023), for example, analyzed 3,071 Science and Mathematics questions from three Indian national school boards administered between 2011 and 2020. Their study examined the extent to which national school-leaving examinations represented lower- and higher-order cognitive levels and demonstrated the continuing importance of Bloom's taxonomy as a framework for evaluating the cognitive demands of large-scale assessments. In the context of assessment design, Scully (2017) similarly emphasized the need to develop examination items capable of measuring higher-order thinking rather than relying exclusively on the recall of factual information.

Taken together, this body of research suggests that the growing emphasis on critical thinking and higher-order cognition in educational policy does not necessarily translate into equivalent changes in examination practices. The persistence of lower-level cognitive items across different educational systems indicates that examination design remains an important area for investigating the alignment between intended learning outcomes and actual assessment practices.

Research has also linked Bloom's higher cognitive levels to broader constructs such as critical thinking (Facione, 1990) and creative problem solving (Lai, 2011). Scholars argue that the top three levels—Analysis, Synthesis/Creation, and Evaluation—map closely onto these constructs, forming the foundation of 21st-century competencies (Anderson & Krathwohl, 2001; Trilling & Fadel, 2009). Consequently, exam systems that omit these levels risk narrowing students' intellectual development and limiting their readiness for complex societal and occupational roles.

2.4. Iranian Research on Cognitive Levels in National Examinations

Research conducted in Iran has similarly identified a predominance of lower-order cognitive processes in formal examinations. Ebadi and Shahbazian (2015) examined final examinations in Iranian high schools using Bloom's taxonomy and provided evidence concerning the cognitive distribution of examination questions in the Iranian context. Their study is particularly relevant to

the present research because it directly addresses the cognitive level of Iranian high-school final examinations.

Further evidence is provided by Rezaee and Golshan (2016), who analyzed nationwide English final examinations administered to Iranian high-school students using Bloom's taxonomy. Their findings showed that the examination questions covered remembering, understanding, and Application, with understanding representing the dominant cognitive level. The study therefore provides direct evidence that Iranian high-stakes language examinations have historically emphasized lower-order cognitive processes.

More recent research indicates that this pattern remains relevant. Ghezelbash, Alizadeh, and Divsar (2025) analyzed English-language tests developed by Iranian high-school teachers across Grades 10 to 12 using a framework incorporating the Revised Bloom's Taxonomy, Webb's Depth of Knowledge, and the Cognitive Rigor Matrix. They found that Remembering, Understanding, and Applying accounted for 89.49% of the test items, whereas Analyzing, Evaluating, and Creating accounted for only 10.51%. These findings provide recent evidence that lower cognitive demand continues to predominate in Iranian secondary-school assessment.

Although these studies differ in subject, examination type, and analytical framework, their findings point toward a recurring pattern: assessment practices in Iran have generally placed greater emphasis on recalling, understanding, and applying established knowledge than on analyzing, evaluating, and creating. This pattern provides an important basis for examining whether the national final examinations analyzed in the present study demonstrate a similar distribution and whether that distribution has changed over time.

3. Methodology

3.1. Research Design

This study employed a quantitative content analysis design with qualitative illustrative examples to examine the cognitive complexity of Iran's Grade-12 national final examinations in two core subjects—Persian Language and Literary Studies and Techniques—across a six-year period (2019–2024). The purpose was to identify patterns and trends in the higher-order cognitive levels of examination questions based on Bloom's taxonomy of the cognitive domain. The research employed quantitative analysis, which focused on the frequency and statistical relationships among cognitive levels and exam years, with qualitative illustrative examples. The study design allowed for an objective assessment of frequency distributions and statistical significance as well as illustrative qualitative examples within the Iranian educational system.

3.2. Sampling

The sample of the study consisted of all national final examinations for Grade 12 in the subjects of Persian Language and Literary Studies and Techniques administered by the Iranian

Ministry of Education between 2019 and 2024. In total, 24 examination papers (6 years $\times$ 2 sessions $\times$ 2 subjects) were analyzed, representing both the Khordad (June) and Dey (January) exam sessions where applicable. The corpus comprised about 800 individual questions, varying in format from multiple-choice and fill-in-the-blank items to short-answer and essay-type prompts. Each question was treated as a separate analytical unit and coded according to Bloom's taxonomy.

3.3. Instrument and Data Coding Procedure

The main analytical instrument was a content-analysis checklist derived from the six cognitive levels of Bloom's revised taxonomy (Anderson & Krathwohl, 2001). Table 1 indicates the checklist which includes operational definitions and keywords for each level.

Table 1

Coding Framework Based on Bloom's Revised Taxonomy

Level	Cognitive Process	Operational Definition / Key Indicators
1	Remember	Recall or recognition of factual information, definitions, or rules
2	Understand	Demonstrate comprehension through explanation, summarization, or paraphrasing
3	Apply	Use learned knowledge or rules to solve familiar problems or perform tasks
4	Analyze	Break information into parts, identify relationships, or infer underlying principles
5	Evaluate	Make judgments based on criteria, compare alternatives, or assess quality
6	Create	Generate new ideas, propose solutions, or integrate elements into novel structures

Following Table 1, each exam question was independently examined and classified into one of the six cognitive categories. Questions involving multiple subcomponents were coded at the highest cognitive level required for a correct response. For example, an item requiring both factual recall and evaluative judgment was coded as "Evaluate." Table 2 shows a sample of questions and their associated categories according to the taxonomy used in the study.

Table 2

A Representative Sample of Questions and Their Related Category According to The Bloom's Taxonomy

Sample question	keywords	Description	Cognitive level
January 2024 - Question 22: Choose the correct arrangement for each verse from the column opposite it. (One arrangement in column "B" is extra) 1- Sensual appeal, 2- Good reasoning, 3- Imagination, 4- Equation style, 5- Paradox	Using, creating, identifying, simulating, applying, etc.	The student must identify the logical structure and cause-and-effect relationship of the text, not simply recall meaning or information.	Analyze
May 2021 - Question 11: In a noun group, use the word horse with a predicate and a postdicate.	Create, combine, propose, create, develop, etc.	The student integrates the knowledge gained and produces something unique.	Create
January 2020 - Question 34: Which characteristics of lovers does the following poem express?	Discover, prove, evaluate, judge, criticize, etc.	The student evaluates and judges the value and validity of various topics.	Evaluate
January 2021 - Question 3: Who is the author of the book "The Grammar of Love?"	Naming, specifying, expressing, praising, etc.	The student remembers the answer and retrieves the answer without change or interpretation.	Remember
May 2023 - Question 8: The meaning of which verse is different from the meaning of "freedom" during the awakening period?	Explain, summarize, recount, describe, etc.	In this section, the student must understand the meaning of the verses and interpret and explain the specified phrases or choose the appropriate option.	Understand

As Table 2 indicates, To better understand the cognitive emphasis of the exams, representative examples were qualitatively examined. In the Remember category, typical questions asked students to “state the definition of metaphor” or “identify the correct grammatical form.” These items required direct recall of factual information without interpretation or application. The Understand level included questions such as “explain the meaning of the underlined verse” or “summarize the main theme of the passage,” which demanded comprehension but not inference.

Questions at the Apply level involved tasks like “use the given rule to complete the sentence correctly” or “apply the principle of meter to the following poem.” These required procedural knowledge but still operated within familiar contexts. A smaller number of Analyze items appeared in literature exams, typically asking students to “compare the imagery of two poets” or “identify the relationship between tone and theme.” Evaluate and Create categories were extremely rare in the sample. Even essay-type prompts tended to restrict responses to predetermined interpretations rather than inviting original judgment or synthesis.

3.4. Reliability and Validity Procedures

To ensure the reliability of the coding process, two trained raters—both experienced educators in Persian Language and Literary Studies and Techniques—independently coded all examination items after receiving instruction on the operational definitions of each cognitive level. The inter-rater reliability coefficient, calculated using Cohen’s Kappa, was 0.87, indicating a high level of agreement (Landis & Koch, 1977). Any disagreements between the raters were discussed and resolved through consensus based on the operational definitions and coding criteria for each cognitive level.

Content validity was established through expert review. Three senior university professors specializing in curriculum and educational assessment evaluated the coding framework for clarity, relevance, and alignment with Bloom’s taxonomy. The experts reviewed the coding framework and its operational definitions; they did not independently recode the examination items. Their feedback was used to confirm the appropriateness and clarity of the coding categories.

3.5. Data Analysis Procedures

During this stage, frequencies and percentages of questions at each cognitive level were calculated for each exam year and for each subject. And in order to test the existence of a statistically significant relationship between exam year and cognitive level, inferential statistics, especially Chi-Square Test of independence was conducted using a 2*2 contingency table. Because some expected frequencies were small, Fisher’s exact test was also considered in interpreting the results. (Field, 2018). The test was computed using SPSS (version 25). Statistical significance was set at $p < 0.05$.

4. Results

The analysis of examination papers from the Persian Language and Literary Studies and Techniques courses over the six-year period (2019–2024) produced a consistent pattern across both subjects: the overwhelming majority of exam questions were concentrated in the lower levels of Bloom’s taxonomy. Across all years, the first three cognitive levels—Remember, Understand, and Apply—accounted for about 80 percent of the total questions analyzed. The upper levels—Analyze, Evaluate, and Create—appeared far less frequently, and the final level (“Create”) was almost absent in all years and both subjects. Table 3 demonstrates distribution of cognitive levels in terms of frequency across years and subjects.

Table 3

Frequency and Percentage Distribution of Cognitive Levels across Years and Subjects

Year	Subject	Remember	Understand	Apply	Analysis	Evaluate	Create
2019-1/2*	Literary	9/7*	4/8	9/6	3/6	0/0	0/0
2019-1/2	Persian L	3/7	26/28	0/3	4/3	0/0	0/0
2020-1/2	Literary	10/11	10/10	8/8	5/7	0/0	0/0
2020-1/2	Persian L	4/8	24/29	0/0	5/3	0/1	1/1
2021-1/2	Literary	10/8	12/11	5/8	4/8	0/0	0/0
2021-1/2	Persian L	5/4	18/17	0/1	6/6	0/1	0/0
2022-1/2	Literary	10/8	10/11	9/6	5/5	0/0	0/0
2022-1/2	Persian L	11/7	24/13	1/1	2/5	0/0	0/1
2023-1/2	Literary	10/7	11/15	1/8	6/6	0/0	0/0
2023-1/2	Persian L	9/8	21/25	0/1	2/4	0/0	0/0
2024-1/2	Literary	8/14	15/11	4/6	7/7	0/0	0/0
2024-1/2	Persian L	9/10	23/21	1/1	3/9	0/0	0/0

*1/2: first / second semester

* Frequency in first/second semester of the mentioned academic year

As Table 3 indicates, descriptive statistics revealed that the Remember category consistently dominated exam content throughout the period under study. During the study period (2019–2024), about 25% of all questions in both subjects required simple recall of memorized facts, definitions, or examples. By contrast, the Understand category represented approximately 50% of items, typically involving explanation or paraphrasing of given texts or grammatical rules. The Apply category accounted for roughly 10% of items, primarily those that asked students to use known rules (e.g., syntax or prosody) in familiar contexts. Figure 1 also represents the yearly distribution of cognitive levels graphically.

Figure 1

Yearly Trend of Cognitive Levels in Persian Language (right figure) and Literary Studies and Techniques (left picture) Exams (2019–2024)

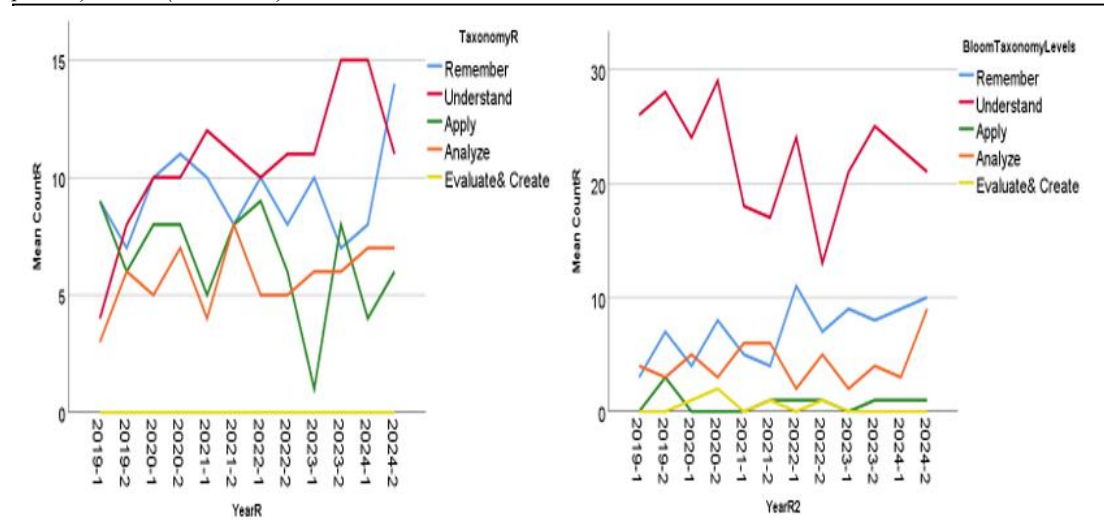


Figure 1 illustrates the yearly distribution of cognitive levels in the Persian Language and Literary Studies and Techniques examinations from 2019 to 2024. Overall, the results show a persistent predominance of lower-order cognitive levels across the examination years, particularly Remembering and Understanding. Although the frequencies of these levels fluctuate from year to year, Understanding generally represents the most frequently observed cognitive level, especially in the right figure. Applying also appears across several examination years but at comparatively lower and more variable frequencies. In contrast, higher-order cognitive levels, particularly Analyzing, remain considerably less frequent, while Evaluating and Creating are largely absent or occur only minimally throughout the period. The year-to-year fluctuations suggest some variation in the cognitive composition of the examinations; however, the overall pattern indicates that the examinations remained substantially concentrated on lower-order cognitive processes, with limited representation of higher-order thinking.

To determine whether there was a significant trend in the presence of higher-order thinking across the examination years, a chi-square test of independence was conducted. First, the examination years were divided into two periods: 2019–2021, coded as 1, and 2022–2024, coded as 2. The examination items were initially classified according to the revised Bloom's taxonomy as either lower-order thinking (Remembering, Understanding, and Applying) or higher-order thinking (Analyzing, Evaluating, and Creating). Given the predominance of lower-order thinking levels, a binary variable was subsequently created to examine the presence of higher-order thinking across the two periods. Items without a higher-order thinking component were coded as 1 (absence), whereas items containing a higher-order thinking component were coded as 2 (presence). A chi-square test of independence was then performed to determine whether the

presence or absence of higher-order thinking was significantly associated with the two examination periods. Table 4 presents the results of this analysis.

Table 4

Chi-square Test of the Association between Examination Period (2019–2021 Vs. 2022–2024) and the Presence or Absence of Higher-Order Thinking Levels

		Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Literary Terms	Pearson Chi-Square	.000 ^a	1	1.000	1.000	.626
	Likelihood Ratio	.000	1	1.000	1.000	.626
	Fisher's Exact Test				1.000	.626
	Linear-by-Linear Association	.000 ^c	1	1.000	1.000	.626
	N of Valid Cases	60				
Literary Studies and Techniques	Pearson Chi-Square	.000 ^a	1	1.000	1.000	.604
	Likelihood Ratio	.000	1	1.000	1.000	.604
	Fisher's Exact Test				1.000	.604
	Linear-by-Linear Association	.000 ^c	1	1.000	1.000	.604
	N of Valid Cases	60				

2*2 Contingency Table : Year (1= 2019-2021; 2=2022-2024)

Bloom Taxonomy 1= Absence of HOT (Analyze, Combine, and Create)

2= Presence of HOT (Analyze, Combine, and Create)

Table 4 presents the results of Chi-Square Test for the association between year and higher-order cognitive levels for both Persian Language and Literary Studies and Techniques. The test didn't show a statistically significant results ($p > 0.05$) for both subjects. Also, the general trend as indicated by charts does not indicate that the distribution of higher-order cognitive levels change across years, suggesting a lack of progressive trend in the distribution of cognitive levels toward higher-order cognitive levels over the years.

5. Discussion

The results of this longitudinal analysis reveal a clear and persistent pattern: Iranian high-school final examinations in Persian Language and Literary Studies and Techniques remain overwhelmingly focused on lower-order cognitive skills. Chi-Square Tests did not confirm statistically significant differences across years and the direction of these differences did not indicate a meaningful or sustained trend toward higher-order cognition. The variations observed appear sporadic, reflecting incidental adjustments in test design rather than systemic reform.

This finding corroborates the conclusions of numerous national and international studies that have documented the dominance of recall-based and comprehension-based questions in standardized examinations (Ebadi & Shahbazian, 2015; Muhayimana et al., 2022; Chowdhury et al., 2023; Ghezelbash et al., 2025). The persistence of this pattern suggests that the Iranian assessment

system—despite explicit curricular emphasis on critical and creative learning—continues to prioritize measurable, objective, and easily scorable outcomes over intellectual depth.

Bloom's taxonomy assumes that effective learning requires progression from simple to complex cognitive operations: learners must first remember information, then understand and apply it, before they can analyze, evaluate, and create. Theoretically, a well-balanced examination should therefore include a distribution of items spanning this continuum, ensuring that students are assessed not only on factual knowledge but also on their ability to interpret, judge, and innovate (Anderson & Krathwohl, 2001).

The near absence of Evaluate and Create questions in the analyzed exams indicates a break in this hierarchy. Without opportunities to demonstrate higher-order skills, students are implicitly encouraged to engage in surface learning. In effect, the examinations appear to measure only the first half of Bloom's hierarchy, thereby narrowing the intended scope of the national curriculum. This imbalance reveals a misalignment between the intended learning outcomes stated in policy documents and the assessed learning outcomes reflected in exam content—a misalignment that undermines constructive alignment (Biggs, 2003).

The findings have significant implications for the development of critical thinking, which is widely recognized as a cornerstone of 21st-century education (Facione, 1990; Lai, 2011). The higher tiers of Bloom's taxonomy—analysis, evaluation, and creation—directly correspond to the cognitive skills involved in critical reasoning and creative problem solving. When exams neglect these levels, they implicitly discourage both teachers and students from engaging in such processes.

In Iran, the Fundamental Reform Document of Education (2011) and the Vision 2025 strategy identify innovation, reasoning, and creativity as essential learning outcomes. However, these objectives remain largely aspirational if assessment practices continue to reward rote memorization. Teachers, constrained by the high stakes of national examinations, are more likely to adopt test-oriented teaching strategies that emphasize recall rather than interpretation or analysis. Consequently, the examination system exerts a conservative influence on classroom practice, perpetuating a cycle of surface learning that contradicts policy objectives.

Beyond the Iranian context, this study contributes to international discourse on how examination systems shape learning cultures. Its findings illustrate the persistence of low-level cognitive assessment even in reform-oriented systems, emphasizing that cognitive transformation requires sustained, multi-layered effort. Bloom's taxonomy remains a valuable diagnostic tool precisely because it exposes the gap between policy rhetoric and assessment reality.

In essence, the discussion underscores that while minor variations in exam composition have occurred over the study period, Iran's national examinations in Persian Language and Literary Studies and Techniques continue to emphasize lower-order thinking. The results confirm that the structural and cultural inertia of standardized testing outweighs reformist intentions. To translate curricular ideals into actual cognitive advancement, systemic redesign of assessment philosophy, question construction, and teacher preparation is required.

This disconnect echoes a common global challenge: large-scale examinations often prioritize standardization, reliability, and cost-efficiency at the expense of validity and depth (Stiggins, 2017). Because subjective or open-ended items are harder to score consistently, exam designers tend to avoid them, even though such items are necessary to assess evaluation and creativity. The Iranian case thus reflects a broader tension between measurement precision and cognitive authenticity in national assessment systems.

6. Conclusion and Recommendations

This study set out to examine the cognitive characteristics of Iranian high-school final examinations in Persian Language and Literary Studies and Techniques over a six-year period (2019–2024) through the lens of the revised Bloom’s taxonomy. By merging and analyzing about 800 examination questions, it sought to identify longitudinal trends in the cognitive demands of these assessments and determine whether a shift toward higher-order thinking has occurred in line with national curriculum reforms.

The study employed a quantitative content analysis with qualitative illustrative examples using Bloom’s revised taxonomy, categorizing each question into one of six cognitive levels: Remember, Understand, Apply, Analyze, Evaluate, and Create.

The findings were unambiguous. Across both subjects and all examined years, lower-order cognitive levels dominated overwhelmingly, accounting for more than 80 percent of all items. Statistical testing did not confirm significant variation in distributions across years and these changes were irregular and non-progressive, showing no sustained movement toward analytical, evaluative, or creative cognition. The uppermost levels of Bloom’s taxonomy were virtually absent, and even when analytical questions appeared, they were constrained by predetermined answers that limited interpretive freedom.

Taken together, these results paint a consistent picture of an assessment system that remains predominantly recall-oriented despite policy commitments to promote creativity and critical thinking. The cognitive profile of the exams does not reflect the learning outcomes envisioned in Iran’s Fundamental Reform Document of Education and related strategic frameworks, which emphasize reasoning, reflection, and innovation. Instead, the persistence of low-level items perpetuates a cycle of rote learning, where teachers teach for memorization and students learn for reproduction rather than understanding.

From an educational perspective, this misalignment between curriculum and assessment undermines the very foundation of reform. Assessment is often described as the “backbone” of education because it shapes how teachers teach and how students learn. If examinations primarily reward recall, they inadvertently discourage analytical thought and problem-solving behavior. Thus, the assessment system, rather than acting as an engine for cognitive development, becomes a constraint on it.

This study contributes to the broader discourse on educational quality by demonstrating how Bloom's taxonomy can function as a diagnostic instrument for evaluating the intellectual rigor of national examinations. The longitudinal and cross-subject approach used here illustrates that the taxonomy is not merely a theoretical model but a practical evaluative tool capable of revealing systemic tendencies over time.

Article in press

References

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
- Biggs, J. (2003). Teaching for quality learning at university (2nd ed.). Open University Press.
- Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). McGraw-Hill Education.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. David McKay.
- Chowdhury, A. R., K. S., Nandagopal, K. S., Prakash, V., & Das, S. (2023). A comparative analysis of the cognitive levels of Science and Mathematics secondary school board examination questions in India. Proceedings of the 16th International Conference on Educational Data Mining.
- 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.
- Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. The California Academic Press.
- Field, A. (2018). Discovering statistics using IBM SPSS Statistics (5th ed.). Sage Publications.
- Forehand, M. (2010). Bloom's taxonomy. In M. Orey (Ed.), Emerging perspectives on learning, teaching, and technology. University of Georgia Press.
- Ghezelbash, A., Alizadeh, I., & Divsar, H. (2025). Discovering levels of cognitive demand in tests developed by Iranian high school English teachers across grades. *International Journal of Contemporary Educational Studies*.
- Gulikers, J. T. M., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcome. *Learning and Instruction*, 18(2), 172–186. <https://doi.org/10.1016/j.learninstruc.2007.02.012>
- Haladyna, T. M. (2004). Developing and validating multiple-choice test items (3rd ed.). Lawrence Erlbaum Associates.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Lai, E. R. (2011). Critical thinking: A literature review. Pearson Education.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lipman, M. (1988). Critical thinking: What can it be?. *Educational Leadership*, 46(1), 38–43.
- Lipman, M. (2003). Thinking in education (2nd ed.). Cambridge University Press.
- Ministry of Education. (2011). Fundamental Reform Document of Education in the Islamic Republic of Iran. Tehran: Supreme Council of Education.

- 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. <https://doi.org/10.1007/s41297-021-00156-2>.
- Rezaee, M., & Golshan, M. (2016). Investigating the cognitive levels of English final exams based on Bloom's taxonomy. *International Journal of Educational Investigations*, 3(4), 57–68.
- Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research & Evaluation*, 22(4), 1–13.
- Stiggins, R. (2017). Revolutionize assessment: Empower students, inspire learning. Corwin Press.
- Supreme Council of Education. (2012). Iran's Vision 2025 educational strategy. Tehran: Ministry of Education.
- Trilling, B., & Fadel, C. (2009). 21st century skills: Learning for life in our times. Jossey-Bass.

Article in press