



AI-GENERATED READING COMPREHENSION QUESTIONS: A COMPARATIVE ANALYSIS THROUGH BLOOM'S AND BARRETT'S TAXONOMIES

Ana Tripković*,
[0009-0009-4541-7380]

Saša Čorboloković
[0009-0002-3767-2313]

Singidunum University,
Belgrade, Serbia

Abstract:

This paper explores the possibilities offered by recent AI models in creating reading comprehension questions. One of the most popular large language models (LLMs), ChatGPT-5.3, was used to generate 15 questions about two short stories, one in Serbian (*Sve će to narod pozlatiti* by Laza Lazarević) and one in English (*The Lottery* by Shirley Jackson). The items were generated for each text using the same prompt in different languages, and subsequently they were analysed through Bloom's revised taxonomy and Barrett's taxonomy of reading comprehension. The findings indicate a predominance of higher-order thinking questions, particularly analysis and evaluation levels, according to Bloom's taxonomy, and inferential comprehension according to Barrett's levels. Both sets of questions show more focus on critical thinking than on simple retrieval of information. One limitation is the lack of tasks at the creative level, indicating that AI-generated questions do not foster productive or generative students' responses. This study suggests that AI has clear potential in language and literature teaching, but also important limitations. It can be concluded that large language models can help teachers by creating useful and intellectually engaging materials, but the materials and tasks produced should not be used without careful review. Teachers still play the most important role in adapting such materials to cover different thinking levels. Overall, these findings are an addition to the current research on AI in education, and they highlight the importance of well-designed prompts in preparing high-quality instructional content.

Keywords:

Reading Comprehension, Bloom's Taxonomy, Barrett's Taxonomy, AI In Education, Literary Text Analysis.

INTRODUCTION

Reading comprehension has continuously evolved, and retrieving information from texts is no longer seen as the most important skill, but real comprehension is a multi-level process involving decoding, interpretation, and meaning construction [1]. Nowadays, reading comprehension includes delving into intricate ideas, using higher cognitive processes to understand what is not explicitly stated. Critical thinking has become the most important skill, and training it through reading comprehension in general education is of paramount importance. Students need to interact more deeply with texts.

Correspondence:

Ana Tripković

e-mail:

atripkovic@singidunum.ac.rs



Using higher-order thinking skills, or HOTs, is the key goal of reading. Learners need to be able to understand structure and subtle ideas, to infer implicit meanings and engage with texts using both cognitive and affective processes. Bloom's taxonomy and Barrett's taxonomy of reading comprehension are two frameworks that are used to determine the levels of thinking necessary to complete certain tasks. They have been used to create classroom materials as well as coursebooks.

The development of large language models (LLMs), as part of generative AI, has opened new perspectives for teachers. They can generate exceptional content within seconds, adapt it to the required levels, create additional tasks, thereby reducing teachers' workload. Nevertheless, the quality of materials created in this way is not always pedagogically aligned or of appropriate quality, which leads to studies such as this one.

The research in this field has focused mostly on the general potential of AI in education, particularly personalisation and efficiency [2, 3]. Less attention has been paid to the cognitive quality of AI-generated tasks, particularly in the domain of reading comprehension. This study addresses the gap by analysing AI-generated material through two different frameworks: Bloom's and Barrett's taxonomies, by comparing well-known English and Serbian literary texts.

2. THEORETICAL FRAMEWORK

2.1. BLOOM'S TAXONOMY

Bloom's taxonomy is one of the most famous frameworks which is used for categorising cognitive processes in education. Originally, it was developed by a Committee of college and university examiners, led by Benjamin Bloom [4], but the version used in most modern pedagogical contexts is a revised version created by Krathwohl [5]. The main idea of this taxonomy was to create a bank of test items, but it was further developed and included in all aspects of education, particularly in curriculum development [6] [7] [8].

According to the original taxonomy, thinking skills are organised in six levels: knowledge, comprehension, application, analysis, synthesis and evaluation. They are placed in a hierarchical order, from lower-order to higher-order thinking. At the beginning of the 21st century, a revised form of taxonomy was proposed in which nouns that represented names of the levels were replaced by verbs and gerunds to highlight the dynamic nature of these levels [9].

The revised list consists of the following levels:

1. Remember (Remembering): retrieving, recognising, and recalling knowledge.
2. Understand (Understanding): interpreting, exemplifying, classifying, summarising, inferring, comparing, and explaining.
3. Apply (Applying): executing, implementing
4. Analyse (Analysing): differentiating, organising, and attributing
5. Evaluate (Evaluating): making judgments, checking and critiquing
6. Create (Creating): generating, planning, or producing. [9]

In addition to changes in naming, the most significant change is the rearrangement of the final levels. Namely, Evaluation was renamed Evaluating and placed in the fifth position instead of the sixth and the highest one it had in the original taxonomy. Additionally, the fifth level from the original, called Synthesis, became the highest level in the revised taxonomy and was renamed Create, which implies putting elements together to create a new whole [10]. The first three levels are considered LOTs (Lower Order Thinking Skills), and the last three levels are HOTs (Higher Order Thinking Skills) [11].

Bloom's taxonomy is valuable for reading comprehension in general, since it distinguishes between levels of cognitive engagement. LOTs such as Remembering and Understanding require students to retrieve explicitly stated information [12], which forms the pedagogical basis of reading comprehension, so these levels need to be included in reading comprehension tasks.

HOTs or Higher-order thinking skills require learners to provide responses on a more complex level. It is the responsibility of a teacher to help students develop HOTs, which also helps them become lifelong learners and is particularly important in language learning [13]. In a literary context, it could include rewriting narratives, adopting different perspectives, expanding topics.

Even though Bloom's taxonomy has faced considerable criticism [14], particularly regarding its linear hierarchy, it is still a useful framework for creating instructions and evaluations, but it might be useful to combine it with other frameworks to achieve its highest potential.



2.2. BARRETT'S TAXONOMY OF READING COMPREHENSION

Barrett's taxonomy is used to assess the cognitive and affective aspects of students' reading comprehension. The main use of this taxonomy is to help teachers construct reading comprehension questions. According to this taxonomy, there are five levels of comprehension: literal comprehension, reorganisation, inferential comprehension, evaluation, and appreciation [15]. This taxonomy differs from Bloom's in the number of levels, but also in the focus of each category. Inferential comprehension, evaluation and appreciation can be correlated with Bloom's higher-order thinking skills (HOTS). Lower-level processes (such as recognising explicitly stated information) are necessary but insufficient for full comprehension. Literal comprehension is not trivial: it is the basis for higher-level thinking [1].

The level that leads to interpretative reading is the inferential level. At this level, meaning is not clearly stated, but has to be constructed from clues in the text and prior knowledge. This level is widely considered central to reading proficiency as it reflects readers' ability to interact with the text [16].

The highest level is appreciation, which differs significantly from Bloom's taxonomy since it includes the affective and aesthetic dimensions of reading [17]. It emphasises emotional response, appreciation for style, imagery, and subjective interpretation. This level distinguishes Barrett's taxonomy from others.

The key strength of Barrett's taxonomy lies in creating reading tasks and focusing on comprehension questions [18], which makes it particularly suitable for adding depth in this type of research. In combination with Bloom's taxonomy, it enables the evaluation of both cognitive levels and interpretative depth.

2.3. AI IN LANGUAGE AND LITERATURE TEACHING

Artificial intelligence in education has changed the approach to teaching and learning. The development of large language models, such as ChatGPT, enabled the generation of texts, questions, lesson plans and tasks, which is beginning to reshape language and literature lessons. Heavy workloads and limited time are constraints that prevent teachers from creating more of their own materials, so they have to use the ones created in coursebooks [19]. This may change with more advanced use of LLMs in pedagogy.

By carefully creating prompts, teachers can generate various levels and tasks to cater to the needs of their students. This includes generating questions that correspond to different levels of Bloom's and Barrett's taxonomies, enabling more targeted and differentiated instruction [20]. Therefore, prompt design becomes crucial, since poorly designed prompts can generate repetitive materials that do not engage higher-order thinking. Moreover, there is a problem with accuracy, bias, and overreliance on these models [21].

Taking all of the aforementioned into account, AI tools should be seen as support for educators, not a replacement. Teachers need to adapt and contextualise materials.

3. METHODOLOGY

3.1. CORPUS

Two short stories were selected as the corpus of this study. For checking reading comprehension in Serbian, the short story by Laza Lazarević, *Sve će to narod pozlatiti* was chosen, whereas the English reading task is *The Lottery* by Shirley Jackson. The texts were selected based on several criteria, with the aims of the research in mind. Firstly, they are both stories that address social themes such as morality, humanity, and human behaviour. Because these topics are central to the stories makes them suitable for interpretation and evaluation at higher levels of thinking. Secondly, the texts are comparable in terms of their length and lexical complexity, thus allowing balanced analysis across the languages. Finally, both works are considered appropriate for 14–15-year-old students, who study English and Serbian as their mother tongue at school. This makes the selection pedagogically relevant and appropriate.

Furthermore, it was decided to use whole stories and not just excerpts, in order to provide a more comprehensive assessment so that the questions could target overall understanding, general interpretation and narrative development.

3.2. PROMPT DESIGN

To make results comparable and consistent, the prompt was carefully designed and an identical prompt was used for both texts, with the only difference being the language used. It follows the principles of effective AI prompting [21] and consists of role assignment, task specification, context and constraints, as well as output formatting. The English prompt was as follows:



You are a language teacher working with students aged 14–15 (native speakers).

Create a list of 15 questions to check comprehension and interpretation of the short story *The Lottery* by Shirley Jackson. The questions should cover different levels of understanding and encourage students to think critically about the text. Format: numbered list.

The prompt was created in this way to intentionally avoid giving additional specifications about the question types, to see what kinds of questions would be generated. Therefore, the prompt has a semi-guided design, where some general guidance was provided, but it leaves enough room for flexibility in output.

3.3. ANALYTICAL PROCEDURE

The questions generated using the prompt were analysed and interpreted based on two different taxonomies, Bloom's revised taxonomy and Barrett's taxonomy of reading comprehension. Each question was carefully, independently examined and assigned to an appropriate category according to both taxonomies. The questions were then compared across the two languages.

4. COMPARATIVE ANALYSIS

Within Bloom's taxonomy, questions were categorised according to six cognitive levels: Remember, Understand, Apply, Analyse, Evaluate, and Create. In parallel, Barrett's taxonomy was used to classify questions into the following categories: literal comprehension, reorganisation, inferential comprehension, evaluation, and appreciation.

4.1. DISTRIBUTION ACCORDING TO BLOOM'S TAXONOMY

The Table 1 below shows the results of the question analysis based on Bloom's revised taxonomy.

The results show that the most dominant cognitive levels are analysis and evaluation with 10 questions each (out of 30). These questions account for 66.6% of all the produced questions. They do belong to higher-order thinking skills, indicating a strong emphasis on these kinds of skills. On the other hand, lower-order processes (Remembering, Understanding, Applying) account for 10 questions (33.3%).

There are no questions in the Create category. There is one borderline question from the English text which could serve as a basis for creative activities (How would the impact of the story change if it were told from Tessie Hutchinson's point of view?), the question itself does not reach the Create level. In order to make it more creative, the task would need to require the students create a new version of the story from a different perspective.

Table 1. Distribution of Questions According to Bloom's Taxonomy

Bloom's level	Serbian text (Lazarević)	English text (Jackson)	Total
Remember	3	2	5
Understand	2	2	4
Apply	1	0	1
Analyse	4	6	10
Evaluate	5	5	10
Create	0	0	0
Total	15	15	30

Table 2. Distribution of Questions According to Barrett's Taxonomy

Barrett's Level	Serbian text (Lazarević)	English text (Jackson)	Total
Literal	4	3	7
Reorganization	1	1	2
Inferential	6	7	13
Evaluation	3	3	6
Appreciation	1	1	2
Total	15	15	30
Total	15	15	30



Bloom's taxonomy analysis reveals that these questions focus on critical thinking and interpretation more than on simple reproduction, but there is no creative production.

Both corpora show a significant number of evaluation questions (5 each), indicating a focus on moral reasoning and critique of society. The Serbian text shows a more even distribution of questions according to Bloom's taxonomy. This is a well-balanced pedagogical progression of questions, definitely really useful for teaching practice.

4.2. DISTRIBUTION ACCORDING TO BARRETT'S TAXONOMY

Table 2 below shows the results of the question analysis based on Barrett's taxonomy for reading comprehension.

The corpora show clear distribution across the levels, with a strong emphasis on the inferential level (43.3%). This suggests that AI can generate questions and focus more on interpretation and implicit meaning, rather than only on explicit meaning.

Literal comprehension questions are important for the basic understanding of a story. From a pedagogical perspective, the lack of such questions would be detrimental to overall understanding, and students need these kinds of questions, particularly at a lower level, so having 23.3% of such questions appears appropriate.

Overall, the two texts have a similar distribution based on Barrett's taxonomy, whereas Bloom's taxonomy shows some variation.

4.3. CROSS-CORPUS COMPARATIVE ANALYSIS

Structurally, both sets emphasise higher-order cognitive skills and show progression from lower-order to higher-order skills. The predominant types of questions in both sets are similar, analysis and evaluation according to Bloom's taxonomy, and inferential comprehension according to Barrett's taxonomy. This indicates that AI-generated questions follow a pedagogical framework in which reading comprehension tasks prioritise interpretation and critical thinking. Another similarity is the lack of the create level of Bloom's taxonomy. This means that students are not required to produce new text or get involved in creative reconstructions. Hence, one of the main limitations of AI in generating questions is the lack of productive or creative output.

A key distinction between the two sets of questions is the type of higher-order thinking they focus on. In the English set of questions, based on the short story *The Lottery*, questions are oriented towards analysis asking students to examine the narrative, interpret certain symbols, identify patterns and relationships and reconstruct meaning. In the Serbian corpus, based on *Sve će to narod pozlatiti*, greater emphasis is placed on evaluation, where questions prompt students to assess the behaviour of certain characters, make moral judgments, evaluate social dynamics, and reflect on broader implications. It is not just about the interpretation of the text, but the interpretation of meaning and values.

In terms of Barrett's taxonomy, both corpora show the dominance of inferential comprehension, again proving that questions generated by AI test more than just explicit content. The difference is that English questions guide students to understand the text, and Serbian questions require students to make judgments about the meaning of the text.

Another feature of the obtained sets is a very limited presence of the reorganisation level (only one in each set), indicating that the traditional comprehension practices seem to be replaced by interpretative tasks.

Reader engagement differs between the two sets. The English corpus focuses more on cognition, interpretation and analysis of the text, in other words, interpretation of textual meaning. The Serbian corpus requires more affective engagement, getting personal responses and focusing on ethics and society through personal experiences and values.

5. DISCUSSION

The comparative analysis shows that both corpora place great emphasis on higher-order thinking and the difference lies in the type of thinking they promote. The English set highlights analytical processing and inferential comprehension, while the Serbian text focuses more on evaluation and more socially grounded interpretation, including moral reasoning. The latter are more reflective and affective. These variations could also be connected to the nature of the texts and their literary discourse.

Furthermore, even though the given sets do emphasise higher-order thinking, there is an obvious gap in creative-level tasks, which points out that even though interpretation and critical thinking are promoted, they remain primarily receptive and interpretative.



Overall, both corpora are aligned with the contemporary approaches to reading comprehension, prioritising critical thinking over superficial understanding. The results contribute to the ongoing debate about the role of AI in language teaching indicating that AI can be a collaborative tool helping educators design effective instructional materials. The generated question sets would need to be critically reviewed and refined by educators, but they can support the development of pedagogically sound materials.

6. CONCLUSION

The main goal of this study was to determine the cognitive quality of AI-generated questions for reading comprehension tasks based on two well-known stories in Serbian and English. The two frameworks were used to draw conclusions: Bloom's taxonomy and Barrett's taxonomy. The findings demonstrate that LLMs do create pedagogically sound questions aligned with the requirements of 21st-century skills. Most questions cover higher-order thinking levels, emphasising skills such as analysis, evaluation and inference.

In both corpora, questions move beyond simple comprehension levels, while still including basic understanding questions to provide a solid foundation. The comparative analysis revealed that the Serbian text focuses more on evaluation and places emphasis on affective level and moral reasoning, whereas the English text included more analytical and inferential questions, aimed at clarifying implicit meanings.

There is a balanced distribution of questions across levels, which is pedagogically sound, and the only real limitation is the lack of the create level according to Bloom. ChatGPT did not require from students to perform productive or transformative analysis of the stories. This indicates the importance of teachers' awareness of these limitations, as well as the advantages and disadvantages of modern tools and their use in pedagogy. LLMs can function as collaborators with educators, refining initial materials provided by AI.

Another aspect worth mentioning is the importance of good prompts and further research could explore different prompt structures within the same context. Moreover, other frameworks could be integrated to assess real-world aspects of reading comprehension, as measured by the PISA reading literacy framework.

REFERENCES

- [1] J. Alderson, *Assessing reading*, Cambridge: Cambridge University Press, 2000.
- [2] A. Fortuna, F. Prasetya, A. D. Samala, S. Rawas, S. Criollo-C, D. Kaya, M. Raihan, W. Andriani, D. Safitri and R. A. Nabawi, "Artificial intelligence in personalized learning: A global systematic review of current advancements and shaping future opportunities," *Social Sciences & Humanities Open*, vol. 12, p. Article 102114, 2025. doi: 10.1016/j.ssa-ho.2025.102114
- [3] A. M. Vieriu and G. Petrea, "The Impact of Artificial Intelligence (AI) on Students'," *Education Sciences*, vol. 15, no. 3, p. 343, 2025. doi: 10.3390/educsci15030343
- [4] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill and D. R. Krathwohl, *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*, B. S. Bloom, Ed., New York: Longmans, Green and co, 1956.
- [5] D. R. Krathwhol, "A Revision of Bloom's Taxonomy: An Overview.," *Theory Into Practice*, vol. 41, no. 4, pp. 212-218, 2002. doi: 10.1207/s15430421tip4104_2
- [6] H. Sönmez, "An Examination of The Revised Bloom's Taxonomy as a Model in the Curriculum Design Process," *Researcher*, vol. 7, no. 1, p. 106-148, 2019.
- [7] N. Doronina, L. Lazorenko and T. Andriichuk, "Enhancing the Practical Learning of English: Applying Bloom's Taxonomy to the ESP Curriculum for Students of Higher Education Institutions," *Economics & Education*, vol. 9, no. 3, pp. 15-20, 2024. doi: 10.30525/2500-946X/2024-3-2
- [8] J. Macalister and I. Nation, *Language Curriculum Design*, New York: Routledge, 2019.
- [9] L. W. Anderson and D. R. Krathwohl, *Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Longman, 2001.
- [10] M. Nikolić and T. Dabić, "The Bloom's Taxonomy Revisited in the Context of Online Tools," in *Sinteza 2016 - International Scientific Conference on ICT and E-Business Related Research*, Belgrade, 2016. doi: 10.15308/Sinteza-2016-315-320
- [11] R. Qasrawi and A. BeniAndelrahman, "The higher and lower-order thinking skills (HOTS and LOTS) in Unlock English textbooks (1st and 2nd editions) based on Bloom's Taxonomy: An analysis study," *International Online Journal of Education and*, vol. 7, no. 3, pp. 744-758, 2020.
- [12] M. Forehand, "Bloom's taxonomy," in *Emerging perspectives on learning, teaching and technology*, M. Orey, Ed., The Global Text Project, 2010, pp. 41-47.



- [13] S. M. Jannah, N. H. P. S. Putro and E. Andriyanti, "Exploring higher order thinking skills (HOTs) through lecturers' beliefs and practices in teaching speaking skills," *Discover Education*, vol. 5, p. Art. no. 113, 2026. doi:10.1007/s44217-026-01233-2
- [14] Ö. G. Ulum, "A Critical Approach to the Inclusion of the Revised Bloom's Taxonomy in Active Skills for Reading: Book 1.," *RumeliDE Dil Ve Edebiyat Araştırmaları Dergisi*, vol. 25, pp. 273-296, 2021. doi: 10.29000/rumelide.1037134
- [15] N. Ramadea, N. Egar and F. Yulianti, "Bloom vs Barret: Comparing reading comprehension types," *Language and Education Journal*, vol. 8, no. 2, p. 128-145, 2023. doi:10.52237/lej.v8i2.486
- [16] W. Grabbe and F. L. Stoller, *Teaching and Researching Reading*, New York: Routledge, 2019.
- [17] E. Junaidi, Y. Zainil and J. Junaidi, "Exploring Effective Reading Tasks for Senior High School: Addressing Key Gaps Using Barrett's Taxonomy," *Elsya : Journal of English Language Studies*, vol. 6, no. 3, pp. 295-314, 2024. <https://doi.org/10.31849/elsya.v6i3.23387>
- [18] D. Köksal, Ö. G. Ulum and N. Yürük, "Revised Bloom's taxonomy in reading texts in EFL/ESL settings," *Acta Educationis Generalis*, vol. 13, no. 1, pp. 133-146, 2023. doi:10.2478/atd-2023-0007
- [19] W. Holmes, M. Bailik and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*, Boston: Center for Curriculum Redesign, 2019.
- [20] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh and S. Günemann, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. Article ID: 102274, 2023. doi:10.1016/j.lindif.2023.102274
- [21] V. Geroimenko, "Key Principles of Good Prompt Design," in *The Essential Guide to Prompt Engineering*. SpringerBriefs in Computer Science, Springer, Cham, 2025, p. 17-36. doi:10.1007/978-3-031-86206-9_2