Psychology, Evaluation, and Technology in Educational Research
8 (2), 2026, 125-136

Available Online: http://petier.org/index.php/PETIER

PETiER
Psychology, Evaluation, and Technology in Educational Research

Is difficult item relevant to higher cognitive process level?: Proof from the geomorphology topic assessment based on Bloom's and Marzano cognitive taxonomy

Listyo Yudha Irawan [1, a] *, Siti Nur Farihah [1, b], Agus Dwi Febrianto [2, c], Aulia Amatullah[1, d], Nurul Ratnawati [1, e], Bayu Kurniawan [1, f]

[1] Universitas Negeri Malang, Semarang St. No.5, Malang, 65145, Indonesia
[2] Universitas Negeri Yogyakarta, Colombo St. No.1, Yogyakarta, 55281, Indonesia
[a] listyo.fis@um.ac.id; [b] siti.nur.1907216@students.um.ac.id; [c] augustd.febrianto@gmail.com;
[d] aulia.amatullah.2207216@students.um.ac.id; [e] bayu.kurniawan.fis@um.ac.id;
[f] nurul.ratnawati.fis@um.ac.id
*Corresponding Author

Received: 19 August 2025; Revised: 27 October 2025; Accepted: 27 April 2026

Abstract: Assessment of learning outcomes needs to consider the appropriateness between the level of difficulty of questions and the cognitive level being measured, especially in Physical Geography–Geomorphology material that requires higher-order thinking skills (HOTS). This study aims to analyze the relationship between the level of difficulty of questions and cognitive levels as defined by Bloom's Taxonomy and Marzano's Taxonomy. The study employs a quantitative descriptive approach by developing ten essay-type questions classified according to both taxonomies. The difficulty level analysis was conducted using Classical Test Theory (CTT) with the parameter of the proportion of participants who answered correctly. The research results indicate that difficulty levels do not always correlate with cognitive levels. In Bloom's Taxonomy, C1 questions tend to be the easiest, while in Marzano's Taxonomy, C3 questions are actually easier than other levels. This demonstrates that difficulty levels and cognitive levels are two distinct concepts, so the perception that HOTS questions are always difficult is not entirely accurate. In conclusion, the development of assessment instruments must consider the relationship between learning objectives, cognitive levels, and item characteristics. The study recommends integrating HOTS-based learning into geomorphology material through the use of contextual media and geospatial technology to support meaningful assessment and the development of 21st-century skills.

Keywords: HOTS Assessment; Item Difficulty; Geomorphology

How to Cite: Irawan, L. Y., Farihah, S. N., Febrianto, A. D., Amatullah, A., Kurniawan, B., & Ratnawati, N. (2026). Is difficult item relevant to higher cognitive process level?: Proof from the geomorphology topic assessment based on Bloom's and Marzano cognitive taxonomy. *Psychology, Evaluation, and Technology in Educational Research*, *8*(2), 125–136. https://doi.org/10.33292/petier.v8i2.290

INTRODUCTION

Learning outcome assessment is conducted to evaluate students' knowledge attainment and competencies as the result of the learning process. Students' learning outcomes can be determined by the instructional programs (Ridha et al., 2025) implemented by teachers in the classroom. For example, in learning models such as problem-based learning and inquiry-based learning, the application of these approaches supports students' learning processes, enabling them to achieve outcomes such as the development of critical thinking skills (Is'ad & Sukarmin, 2022; Sutiani et al., 2021). Furthermore, specific skills or competencies can also be fostered through other instructional models, such as creative thinking through project-based learning (Biazus & Mahtari, 2022), collaboration skills through cooperative learning, and spatial thinking skills through spatial-based learning (Manek, 2023)

It is challenging to ascertain whether a student possesses higher-order thinking skills (HOTS) if the learning process they experience does not facilitate such cognitive development. The cultivation of HOTS is generally fostered through learning activities such as student-centered learning and active learning (Akyol & Garrison, 2011; Limbach & Waugh, 2010; Retnawati et al., 2018). In this regard, it is essential to evaluate the implementation of instructional programs to ensure that students' knowledge and competencies result from meaningful learning experiences (Umami, 2018). Learning that fails to support significant knowledge or competency development risks becoming meaningless, resulting in no substantial improvement or enrichment of students' abilities, which in turn may lower students' motivation to engage in learning (Agra et al., 2019)

The quality of learning assessment outcomes can depend on several factors, one of which is the quality of the test as an instrument for revealing students' knowledge, competencies, and abilities after participating in a learning program (Tiban et al., 2021). An effective test must demonstrate high measurement validity, meaning it should accurately assess the intended constructs (Azwar, 2018). In addition, item characteristics play a critical role in influencing the accuracy of learning outcome measurements. Among these, item difficulty refers to the probability that a test item will be answered correctly or incorrectly by students, while item discrimination reflects the ability of an item to differentiate between high-achieving and low-achieving students (Daryanto, 2010; Kaka et al., 2024).

Recent trends underscore the necessity for education to develop individual potential and skills across diverse subjects and disciplines. Twenty-first-century skills include cognitive abilities and competencies that require enhancement. Consequently, relevant instruments are needed to assess abilities across a range of cognitive levels, from higher-order thinking skills (HOTS) to lower-order thinking skills (LOTS). One approach to gauging the complexity of cognitive abilities is through Bloom's and Marzano's taxonomies, both of which classify cognitive levels according to the complexity of thinking processes. These frameworks enable the development of assessment instruments containing items that measure both HOTS and LOTS.

In practice, Bloom's taxonomy comprises six cognitive levels, ranging from C1 to C6. Levels C1 to C3 are categorized as lower-order thinking skills (LOTS), while levels C4 to C6 are associated with higher-order thinking skills (HOTS). Meanwhile, Marzano's cognitive taxonomy represents a multidimensional framework encompassing distinct dofmains, namely the metacognitive system, the cognitive system, the self-system, and knowledge (Sun et al., 2023). Accordingly, both Bloom's and Marzano's taxonomies adopt different approaches in conceptualizing cognitive abilities. Nevertheless, in educational practice, these taxonomic frameworks are frequently employed as essential guidelines for determining the complexity of cognitive processes in learning.

Instruction that supports both simple (lower-order) and complex (higher-order) thinking should be applied across all subjects, including Physical Geography, specifically Geomorphology. Students should not only recall or understand the causes and effects of studied phenomena but also be able to propose ideas for addressing or solving problems related to the subject matter. Therefore, assessments should consider multiple cognitive levels to ensure comprehensive evaluation and to encourage the development of HOTS, which are essential for addressing real-world challenges.

The development of learning outcome tests often relies on cognitive level and item difficulty as guiding principles. However, there is a tendency to assume that items measuring higher cognitive levels must be difficult, and vice versa that items measuring lower cognitive levels must be easy. In reality, cognitive level and difficulty are distinct concepts: cognitive level refers to the mental effort required to solve an item, whereas difficulty level refers to the proportion of students answering the item correctly or incorrectly (Rajagukguk & Naibaho, 2023).

Despite extensive theoretical discussion on Bloom's and Marzano's Taxonomies, empirical research directly examining the correspondence between item difficulty and cognitive levels remains scarce. Previous studies have mainly focused on developing HOTS-based assesments without analyzing whether higher-order questions are indeed more difficult in practice. Furthermore, there is a lack of evidence from Physical Geography, particularly Geomorphology, where conceptual understanding and analytical reasoning are equally critical.

This study aims to analyze test items measuring students' learning outcomes in the topic of Physical Geography Geomorphology, focusing on the evaluation of item difficulty in relation to cognitive level as reflected in Bloom's and Marzano's taxonomies.

METHODS

Focus Study

The research conducted using a descriptive analysis design with a quantitative approach. The primary focus was to explore the alignment between item difficulty levels and the cognitive level of the test items developed. The construction of test items was based on the cognitive ability levels outlined in Bloom's Taxonomy and Marzano's Taxonomy. Consequently, each test item was categorized according to both Bloom's and Marzano's cognitive level classifications.

Subsequently, the study proceeded with an item analysis. In this context, each test item could be assigned to different taxonomy levels due to the application of two distinct taxonomies in item construction. However, each item possessed only one difficulty index value, determined through the item analysis process. This allowed for the assessment of the consistency between item difficulty levels by comparing the classification of items according to Bloom's Taxonomy and Marzano's Taxonomy. The overall research framework is illustrated in the following diagram (Figure1).

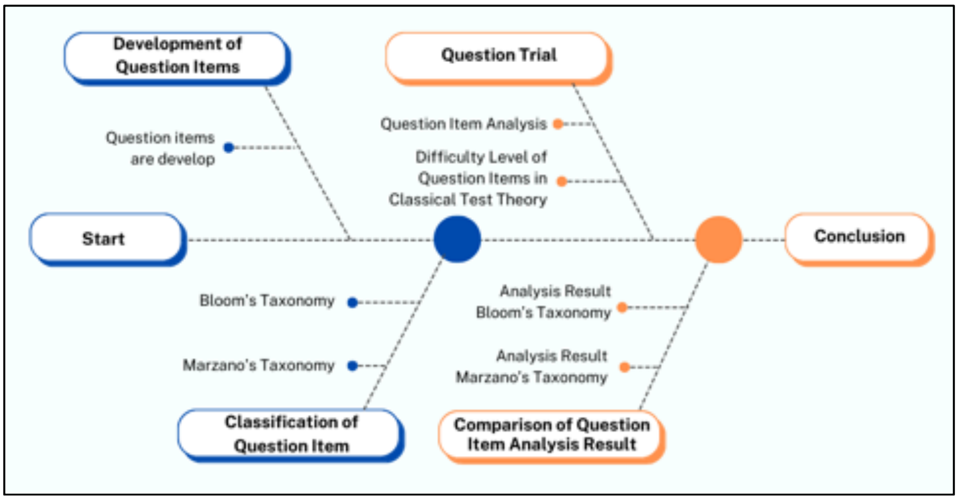

Figure 1. Research Flowchart

Bloom and Marzano Cognitive Taxonomy

Taxonomies of Ability represent conceptual frameworks that describe the hierarchy or levels of proficiency or competence of individuals across various domains. Prominent educational theorists such as Bloom and Marzano have each proposed their own taxonomies of individual abilities, grounded in their respective theoretical considerations. Bloom, for instance, conceptualized cognitive abilities in a structured hierarchy ranging from the lowest to the highest levels (Ulfah & Arifudin, 2023). Over time, Bloom's Taxonomy has frequently been employed as a foundation for designing both learning objectives and assessment frameworks (Nafiati, 2021). Through the use of Bloom's Taxonomy, assessment and evaluation can be designed to align integrally with predetermined learning objectives and instructional programs. Bloom's framework also classifies abilities acquired through learning into three domains: cognitive, psychomotor, and affective.

In contrast, Marzano proposed a more complex multidimensional representation of individual abilities. According to Marzano, the outcomes of learning are not limited to the stage at which individuals acquire or manipulate knowledge (Suhendro et al., 2021). Instead, they extend to considerations of relevance, urgency, and self-evaluation in the acquisition of that knowledge. Learning outcomes, therefore, are influenced not only by cognitive processes but also by additional factors such as emotion, motivation, and an individual's self-awareness in managing their own learning (Azzahra et al., 2025).

A commonality between Bloom's Taxonomy (BT) and Marzano's Taxonomy (MT) lies in their hierarchical structuring of knowledge based on learning outcomes, ranging from lower-order thinking skills (LOTS) to higher-order thinking skills (HOTS). In Bloom's hierarchy, the levels progress from Remembering (C1), Understanding (C2), Applying (C3), Analyzing (C4), Evaluating (C5), to Creating (C6) (Ramdhani et al., 2024). Meanwhile, Marzano's hierarchy ranges from Level 1 (Retrieval) and Level 2 (Comprehension) to Level 3 (Analysis) and Level 4 (Knowledge Utilization) (Saputri et al., 2019). Given these clear frameworks for categorizing cognitive levels, the present study evaluates the outcomes of test item development by examining the consistency of item parameter values in relation to the cognitive levels defined in each taxonomy.

Question Quality and Spesifications Analysis

The quality of a test can be evaluated through item parameters, including difficulty level and discrimination index (Erfan et al., 2020; Ramadhan & Hidayatullah, 2023). Another parameter, guessing, is specifically used to assess the quality of multiple-choice items. The difficulty level refers to the distribution of individuals (Nuridayanti et al., 2024) who answer an item incorrectly (Son, 2019), while the discrimination index indicates the item's ability to distinguish between individuals with high and low levels of ability (Setiyawan & Wijayanti, 2020; Saputri et al., 2023).

Identifying item parameters is essential in test development, as high-quality items are crucial for the assessment and evaluation of instructional programs by teachers or instructors. At the higher education level, test items are often designed to measure higher-order thinking skills (HOTS), with the aim of supporting the evaluation and enhancement of competencies relevant to specific fields of study.

The items developed in this study are in the form of essay questions. In general, essay questions are considered more effective than multiple-choice questions for measuring HOTS (Inayati et al., 2024; Hardina, 2025). Essays allow test-takers to express a broad range of ideas related to the question, whereas multiple-choice items restrict responses to predetermined options and are therefore less able to capture relevant but unconventional answers.

The item analysis in this study is limited to the difficulty parameter, using the Classical Test Theory (CTT) approach. The difficulty parameter is often associated with the cognitive taxonomy applied in test development. While high difficulty levels may suggest that individuals who answer the item correctly possess high ability, this does not necessarily indicate that they have advanced cognitive skills. Difficulty level and cognitive taxonomy are distinct components: difficulty is an inherent attribute of the test item, while higher-order thinking skills (HOTS) are characteristics of the individuals being assessed.

In classical test theory, this is represented by the following equation.

$$P(\%) = \frac{Number\ of\ testees\ who\ answered\ correctly}{Total\ number\ of\ testee} \quad (1)$$

with interpretation

$P \geq 0.70$: Easy question (≥ 70% participants answered correctly)
$0.30 \leq P < 0.70$: Medium difficulty questions (30%–69% participants answered correcty)
$P < 0.30$: Difficult question (< 30% participants answered correctly)

RESULTS AND DISCUSSION

Differences in Levels of Difficulty in Bloom's Taxonomy and Marzano's Taxonomy

The items developed in this study are essay-type questions. The scoring was conducted using a classification scale ranging from 1 to 4, where a score of 1 represents the lowest achievement and a score of 4 represents the highest. A total of 10 items were developed, each of which was classified according to its cognitive level based on both Bloom's Taxonomy and Marzano's Taxonomy. For items categorized at the same cognitive level, the difficulty index was calculated as the average of the difficulty values for those items. Consequently, each cognitive level was assigned a single difficulty index value. The results of the item difficulty analysis are presented in the following table.

Table 1. Distribution of Question Difficulty Levels in Each Taxonomy

Score	Bloom's Taxonomy			Marzano's Taxonomy		
	C2	C3	C4	C1	C2	C3
1	0.05	0.07	0.20	0.03	0.07	0.11
2	0.19	0.12	0.28	0.06	0.22	0.21
3	0.39	0.32	0.27	0.25	0.45	0.31
4	0.37	0.49	0.24	0.66	0.27	0.37
Avg	0.25	0.25	0.25	0.25	0.25	0.25
Std.Dev	0.14	0.17	0.03	0.25	0.14	0.10

The score distribution, as presented in the table above, illustrates the difficulty level of the developed essay items. In essay-type assessments, difficulty level can be determined based on the distribution of test-takers across score categories. A score of 4 represents the highest achievement category; therefore, if the majority of test-takers obtain a score in this category, the item can be classified as "easy." Conversely, if the majority of test-takers score in the lowest category (score 1), the item can be considered "difficult."

Items categorized at the cognitive level C2 in Bloom's Taxonomy or C1 in Marzano's Taxonomy fall within the "moderate" difficulty category. This classification is based on the score distribution, where more than 30% of test-takers obtained scores in categories 3 and 4 (for Bloom's Taxonomy), and more than 60% obtained scores in category 4 (for Marzano's Taxonomy).

Items at the C2 (Understanding) level represent the lowest cognitive level in Bloom's Taxonomy, while items at the C1–Level 1 (Retrieval) level represent the lowest cognitive level in Marzano's Taxonomy. The highest cognitive level items in each taxonomy are C4 (Analyzing) in Bloom's Taxonomy and C3–Level 3 (Analyzing) in Marzano's Taxonomy. The data analysis results do not indicate high difficulty levels for items at the highest cognitive levels. This is evidenced by the increasing number of individuals in higher score categories (C3 – Bloom's Taxonomy). Overall, the distribution of item difficulty levels for each taxonomy is presented in Figures 2 and 3.

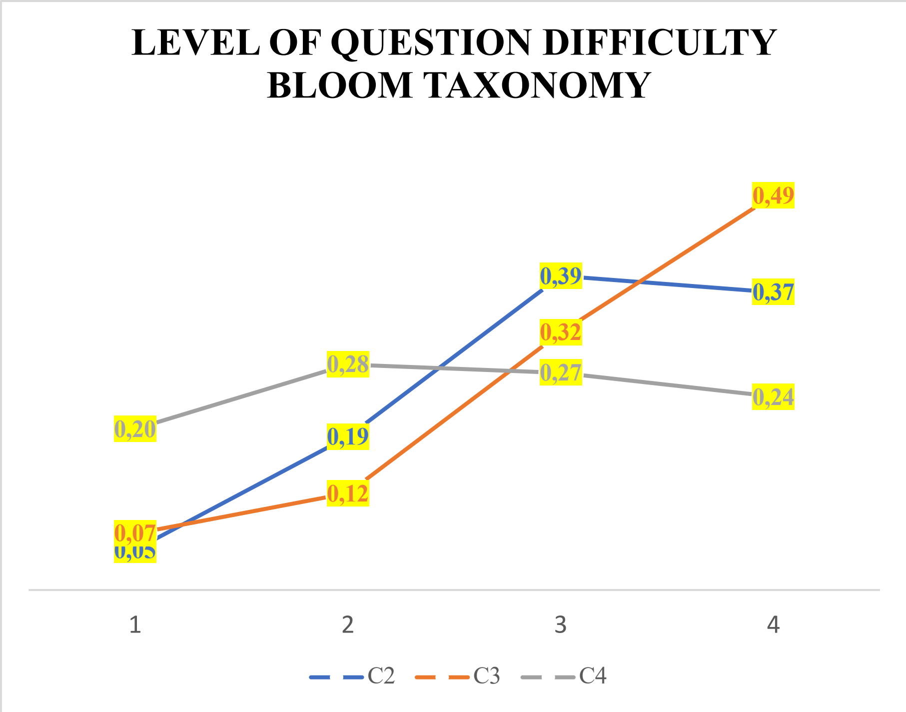

Figure 2. Graph of Difficulty Level Distribution according to Bloom's Taxonomy

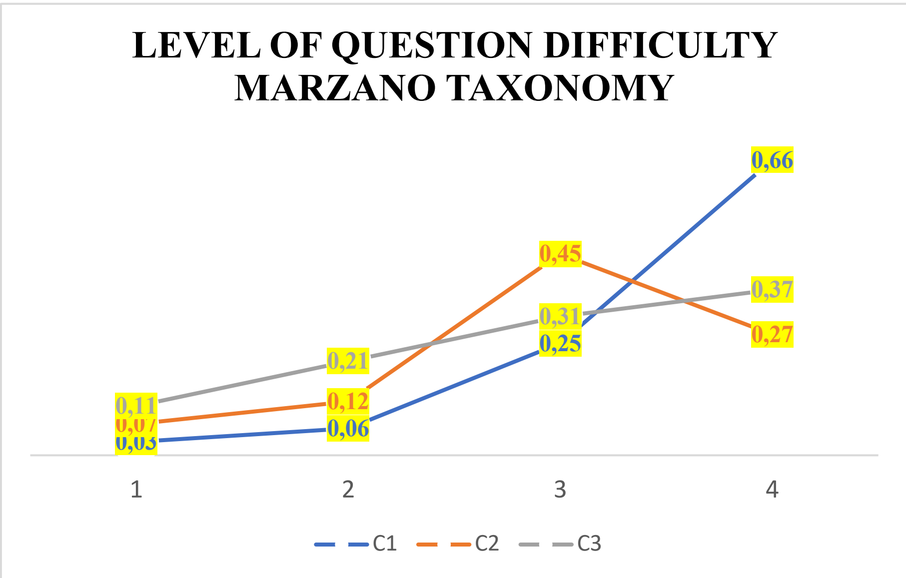

Figure 3. Graph of Difficulty Level Distribution according to Marzano's Taxonomy

In general, both in Bloom's and Marzano's cognitive taxonomies, no items were found with a difficulty index greater than 50%, meaning that only a small proportion less than half of the test-takers obtained a score in category 1. Furthermore, the distribution of test-takers across score categories did not consistently increase in line with the rise in cognitive level. For example, in Bloom's Taxonomy, the lowest-level items (C1, blue line) were the easiest, whereas in Marzano's Taxonomy, the easiest items were those at level C3 (orange line).

The standard deviation values indicate a tendency for the number of test-takers in each score category to become more evenly distributed as the cognitive level increases. In other words, the higher the cognitive level, the more balanced the distribution of scores across categories 1, 2, 3, and 4, with minimal variation in the number of test-takers in each category. Further research is required to determine the significance of this finding, and thus it cannot yet be concluded that higher cognitive levels necessarily correspond to greater item difficulty for test-takers.

Compilation of Physical Geography Questions - Geomorphology

Higher Order Thinking Skills (HOTS)-based questions aim to develop students' critical (Laksono, 2018), analytical, creative, and evaluative thinking abilities. In constructing science-related test items, particularly in physical geography topics such as geomorphology, the questions should incorporate elements that stimulate students to comprehend, apply, analyze, evaluate, and devise solutions to phenomena or problems related to landforms on the Earth's surface, which constitute the core focus of geomorphological studies.

Geomorphology encompasses the study of various landforms, including those shaped by endogenic processes (structural, volcanic) and exogenic processes (from denudational to glacial) (Kurnianto, 2019). Such a scope inherently requires supportive factors, such as providing relevant visual stimuli. Several principles are essential in developing HOTS-based assessment items in geomorphology. One of the key requirements is the use of contextual and engaging stimuli (Fanani, 2018; Suhendro et al., 2021), such as geomorphological maps, landscape photographs, graphs, data tables, or descriptive accounts of natural phenomena; and (2) designing questions that enable students to solve problems by enhancing their analytical and synthetic skills, for example, by linking geomorphological concepts to the natural processes that shape Earth's landscapes.

HOTS-based questions also serve as tools for evaluating the effectiveness of teaching and learning. In other words, HOTS cannot be adequately measured even with well-developed items if instruction does not actively train higher-order thinking processes. Such questions are only valid if preceded by learning activities that stimulate higher-order thinking.

For example, the question "Explain the differences and similarities in landscape characteristics based on the influence of endogenic and exogenic processes!" requires students to identify two main factors (endogenic and exogenic forces), analyze their interrelationships, and compare the characteristics of the resulting landforms. This task falls under the C4 (analysis) category in Bloom's taxonomy, demanding students to differentiate concepts and make comparisons grounded in scientific arguments.

Conversely, the question "Explain the stages of erosion in the formation and development of landforms!" is considered easier, as the instructions are explicit and the scope is specific. It belongs to the C2 (comprehension) level, where students are merely expected to describe the stages without engaging in complex reasoning or argumentation, making it more reproductive than reflective.

Another example, "Give three examples demonstrating that volcanic processes are geomorphic agents capable of changing landscapes!", requires students to evaluate the role of volcanic processes as geomorphic agents and provide three relevant examples, along with logical explanations of how these processes contribute to landform changes. This question falls under C5 (evaluation), as it demands judgment and justification supported by scientific evidence.

HOTS-based assessment in physical geography, particularly in fundamental geomorphology, should not merely test students' theoretical understanding. It must also evaluate their ability to connect concepts with real-world contexts, conduct critical analyses, and develop creative solutions to complex problems. Such an approach fosters the higher-order thinking skills essential for comprehensively understanding the dynamics of geomorphological processes.

The Urgency of Developing HOTS in Physical Geography Learning

Higher Order Thinking Skills (HOTS) refer to cognitive processes that go beyond the mere memorization of theories (Yen & Halili, 2015; Widyawati et al., 2021). Instead, HOTS emphasize the ability to reinterpret and communicate acquired information effectively to others (Astuti, 2018). HOTS require individuals to engage in active cognitive processing when confronted with specific factual situations. In the context of physical geography, problems are often framed in relation to everyday and factual natural phenomena, such as volcanic eruptions, which are caused by endogenous forces, the impact of such phenomena, and the appropriate mitigation measures to minimize the potential impact of disasters (Irawan et al., 2024). Students are expected to explain the role of endogenous forces in the occurrence of volcanic eruptions within their locality. This demands more than basic comprehension; it requires the ability to critically assess, draw logical conclusions, and propose effective solutions in response to the information or phenomena presented.

Physical geography learning particularly in geomorphology based on HOTS plays a crucial role in fostering students' critical and analytical thinking skills. The integration of HOTS into geography instruction enables students not only to grasp fundamental concepts but also to analyze, evaluate, and devise solutions to complex geomorphological problems. HOTS-based geography assessment encourages students to think critically and creatively in understanding geosphere phenomena (HSB & Mukminan, 2024), thereby equipping them with essential higher-order thinking capabilities to address global challenges.

The implementation of HOTS-based geomorphology learning plays a significant role in enhancing students' problem-solving and decision-making skills. Through this approach, students are guided to identify, analyze, and formulate solutions to various geomorphological issues, such as erosion, landslides, and landform changes. Geography education, therefore, is directed toward fostering higher-order thinking, critical reasoning, and creativity. Consequently, students acquire not only theoretical knowledge but also contextual understanding that reflects current environmental conditions, enabling them to fully comprehend geomorphological phenomena, particularly those occurring in their immediate surroundings (Irawan et al., 2022).

Moreover, HOTS-based geomorphology learning promotes the integration of diverse information sources and technologies into the learning process. The use of tools such as Geographic Information Systems (GIS) and geospatial data allows students to conduct in-depth spatial analyses, thereby deepening their understanding of Earth's surface dynamics. As emphasized by Rizqi and Rahmah (2023), geography assessments in the 21st century must incorporate higher-order thinking skills that align with advancements in technology and information. In this way, students are prepared to become adaptive and competent individuals capable of leveraging technology to solve physical geography problems.

CONCLUSION

The study concludes that item difficulty and cognitive level are two independent constructs in the assessment of learning outcomes. While item difficulty reflects how many students can correctly answer a question, cognitive level represents the mental process involved in solving it. The results show that questions categorized as higher order thinking skills (HOTS) are not always more difficult than those in lower order thinking skills (LOTS), indicating that difficulty does not directly correspond to cognitive complexity. Therefore, in Physical Geography particularly in Geomorphology developing valid, problems-based assessment instruments is essential for accurately measuring students high order thinking skills abilities.

REFERENCES

Agra, G., Formiga, N. S., de Oliveira, P. S., Costa, M. M. L., Fernandes, M. D. G. M., & da Nóbrega, M. M. L. (2019). Analysis of the concept of meaningful learning in light of the Ausubel's theory. *Revista Brasileira de Enfermagem, 72*(1), 248–255. https://doi.org/10.1590/0034-7167-2017-0691

Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology, 42*(2), 233–250. https://doi.org/10.1111/j.1467-8535.2009.01029.x

Astuti, P. (2018). Kemampuan literasi matematika dan kemampuan berpikir tingkat tinggi. *PRISMA: Prosiding Seminar Nasional Matematika, 1*, 263–268. https://journal.unnes.ac.id/sju/index.php/prisma/

Azwar, S. (2018). *Metode penelitian psikologi* (Edisi II). Pustaka Pelajar. http://pustakapelajar.co.id/buku/metode-penelitian-psikologi-ed-2/

Azzahra, D., Dewi, P., Mutia, N., & Asna, A. (2025). Belajar dan perilaku belajar. *Formatif: Jurnal Pendidikan Sosial dan Humaniora, 1*(2), 58–69. https://glonus.org/index.php/formatif/article/view/158/123

Biazus, M. de O., & Mahtari, S. (2022). The impact of project-based learning (PjBL) model on secondary students' creative thinking skills. *International Journal of Essential Competencies in Education, 1*(1), 38–48. https://doi.org/10.36312/ijece.v1i1.752

Daryanto. (2010). *Evaluasi pendidikan* (6th ed.). Rineka Cipta. https://perpustakaan.binadarma.ac.id/opac/detail-opac?id=3754

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model Rasch. *Indonesian Journal of Educational Research and Review, 3*(1), 11. https://doi.org/10.23887/ijerr.v3i1.24080

Fanani, M. (2018). Strategi pengembangan soal HOTS pada Kurikulum 2013. *EDUDEENA, 2*(1). https://doi.org/10.30762/ed.v2i1.582

Hardina. (2025). *Analisis soal higher order thinking skills (HOTS) pada buku Cakap Berbahasa dan Bersastra Indonesia kelas XI SMA* [Skripsi, Universitas Negeri Makassar]. https://lib.unm.ac.id/storage/file_thesis/bjtKuxwedQB5y62TdUXqIx5TNeRwA76afM7D1KlM.pdf

HSB, T. S., & Mukminan, M. (2024). Development of hydrosphere geography learning media to improve higher order thinking skills (HOTS). *International Journal of Humanities Education and Social Sciences, 4*(3). https://doi.org/10.55227/ijhess.v4i3.1411

Inayati, N. L., Fatimah, A. N., Azzahra, S. E., & Alamsyah, I. R. (2024). Implementasi tes essay dalam evaluasi pembelajaran pendidikan agama Islam. *Khatulistiwa: Jurnal Pendidikan dan Sosial Humaniora, 4*(1), 114–120. https://doi.org/10.55606/khatulistiwa.v4i1.2724

Irawan, L. Y., Farihah, S. N., Prasetyo, W. E., Azizah, V., Hartono, R., Masruroh, H., & Mapa, M. T. (2024). The effectivity of community-based early warning system on reducing Semeru eruption impact post-eruption 2021. *IOP Conference Series: Earth and Environmental Science, 1314*(1). https://doi.org/10.1088/1755-1315/1314/1/012037

Irawan, L. Y., Sumarmi, S., Febrianto, A. D., Sholeha, A. W., Dirgantara, S. D., Nadhifah, M. R., Wahono, F., & Panoto, D. (2022). Local material in improving students' knowledge and comprehension of natural hazards. In *Exploring new horizons and challenges for social studies in a new normal* (pp. 88–93). Routledge. https://doi.org/10.1201/9781003290865-17

Is'ad, N., & Sukarmin, S. (2022). Implementation of problem-solving learning model assisted by student worksheets to improve critical thinking skills in the context of reaction rate. *Jurnal Pijar Mipa, 17*(2), 199–208. https://doi.org/10.29303/jpm.v17i2.3296

Kaka, L., Bano, V. O., & Njoeroemana, Y. (2024). Efektivitas analisis butir soal pilihan ganda menggunakan aplikasi Anates di SMPN 2 Kanatang. *Jurnal Inovasi Penelitian, 4*(9), 1441–1450. https://doi.org/10.47492/jip.v4i9.3124

Kurnianto, F. A. (2019). Proses geomorfologi dan kaitannya dengan tipologi wilayah. *Majalah Pembelajaran Geografi, 2*(2), 131–147. https://mpg.jurnal.unej.ac.id/index.php/PGEO/article/download/13538/7097/28164

Laksono, P. J. (2018). Pengembangan dan penggunaan instrumen two-tier multiple choice pada materi termokimia untuk mengukur kemampuan berpikir kritis. *Orbital: Jurnal Pendidikan Kimia, 2*(2), 80–92. https://doi.org/10.19109/ojpk.v2i2.2646

Limbach, B., & Waugh, W. (2010). Developing higher level thinking. *Journal of Instructional Pedagogies, 9*. https://files.eric.ed.gov/fulltext/EJ1097083.pdf

Manek, A. H. (2023). Pengaruh model spatial based learning (SBL) terhadap kemampuan berpikir spasial mahasiswa pada mata kuliah Praktek Kerja Lapangan Geografi (PKLG). *Jurnal Geoedusains, 4*(1). https://doi.org/10.30872/geoedusains.v4i1.2131

Nafiati, D. A. (2021). Revisi taksonomi Bloom: Kognitif, afektif, dan psikomotorik. *Humanika, 21*(2), 151–172. https://doi.org/10.21831/hum.v21i2.29252

Nuridayanti, Arthur, R., Rahayu, W., & Ishak, A. M. (2024). Analisis butir soal ujian tengah semester mata kuliah sistem digital dengan pendekatan Rasch model. *INTEC Journal: Information, Technology, Education Journal, 3*(3), 36–42. http://journal.unm.ac.id/index.php/INTEC/article/view/5376/3394

Rajagukguk, M. J. T., & Naibaho, D. (2023). Mampu memilih soal berdasarkan tingkat kesukaran. *Pediaqu: Jurnal Pendidikan, Sosial dan Humaniora, 2*(4), 12736–12747. https://publisherqu.com/index.php/pediaqu/article/view/701/645

Ramadhan, A. F., & Hidayatullah, R. S. (2023). Analisis kualitas butir soal ujian satuan pendidikan (USP) materi C2 teknik pemesinan kelas XII di SMK PGRI 1 Lamongan melalui model Rasch. *Jurnal Pendidikan Teknik Mesin, 12*(3), 1–10. https://ejournal.unesa.ac.id/index.php/jurnal-pendidikan-teknik-mesin/article/view/56146

Ramdhani, S. S., Susanti, R., & Meilinda. (2024). Cognitive level of Program for International Student Assessment (PISA) questions based on the revised Bloom's taxonomy. *European Journal of Education and Pedagogy, 5*(2), 104–112. https://doi.org/10.24018/ejedu.2024.5.2.785

Retnawati, H., Djidu, H., Apino, E., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st Century, 76*(2), 215–230. https://doi.org/10.33225/pec/18.76.215

Ridha, A. R., Rahmatullah, A., Nisrina, A., Firdaus, N., & Wicaksono, Y. (2025). Hasil belajar sebagai objek penilaian (kognitif, afektif, dan psikomotorik). *INOVASI: Jurnal Ilmiah Pengembangan Pendidikan, 3*(3), 1–8. https://ejournal.lpipb.com/index.php/inovasi

Rizqi, I., & Rahmah, N. (2023). Peningkatan keterampilan geografis peserta didik di pembelajaran abad 21. *Geography Science Education Journal (GEOSEE), 4*. http://jurnal.unsil.ac.id/index.php/geosee/index

Saputri, A. F., Susilo, D. A., & Fayeldi, T. (2019). Kemampuan pemecahan masalah peserta didik menyelesaikan soal cerita ditinjau dari taksonomi Marzano. *Semnas SENASTEK Unikama 2019, 2*. http://conference.unikama.ac.id/artikel/index.php/senastek/article/view/164/139

Saputri, H. A., Zulhijrah, Larasati, N. J., & Shaleh. (2023). Analisis instrumen asesmen: Validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri, 9*(5). https://doi.org/10.36989/didaktik.v9i5.2268

Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis kualitas instrumen untuk mengukur kemampuan pemecahan masalah siswa selama pembelajaran daring di masa pandemi. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika, 1*(2), 130–139. https://doi.org/10.46306/lb.v1i2

Son, A. L. (2019). Instrumentasi pemecahan masalah matematis: Analisis reliabilitas, validitas, tingkat kesukaran dan daya beda butir soal. *Gema Wiralodra, 10*(1), 41–52. https://doi.org/10.31943/gemawiralodra.v10i1.8

Suhendro, Sugandi, D., & Ruhimat, M. (2021). Analysis on factors influencing geography teachers' ability in constructing high-order thinking skills (HOTS) assessment instrument. *Geosfera Indonesia, 6*(2), 205. https://doi.org/10.19184/geosi.v6i2.21428

Sun, L., Hu, L., Zhou, D., & Yang, W. (2023). Evaluation and developmental suggestions on undergraduates' computational thinking: A theoretical framework guided by Marzano's new taxonomy. *Interactive Learning Environments, 31*(10), 6588–6610. https://doi.org/10.1080/10494820.2022.2042311

Sutiani, A., Situmorang, M., & Silalahi, A. (2021). Implementation of an inquiry learning model with science literacy to improve student critical thinking skills. *International Journal of Instruction, 14*(2), 117–138. https://doi.org/10.29333/iji.2021.1428a

Tiban, F. A., Achmad, N., & Zakiyah, S. (2021). Analisis kualitas tes hasil belajar matematika buatan guru di SMP Negeri 4 Gorontalo. *Jambura Journal of Mathematics Education, 2*(2), 70–77. https://doi.org/10.34312/jmathedu.v2i2.11103

Ulfah, U., & Arifudin, O. (2023). Analisis teori taksonomi Bloom pada pendidikan di Indonesia. *Jurnal Al-Amar: Ekonomi Syariah, Perbankan Syariah, Agama Islam, Manajemen dan Pendidikan, 4*(1), 13–22. https://ojs-steialamar.org/index.php/JAA/article/view/87

Umami, M. (2018). Penilaian autentik pembelajaran Pendidikan Agama Islam dan Budi Pekerti dalam Kurikulum 2013. *Jurnal Kependidikan, 6*(2), 222–232. https://doi.org/10.24090/jk.v6i2.2259

Widyawati, A., Siti, I. A., & Dwiningrum, R. (2021). Pembelajaran ethnosciences di era revolusi industri 4.0 sebagai pemacu higher order thinking skills (HOTS). *Jurnal Pembangunan Pendidikan: Fondasi dan Aplikasi, 9*(1), 66–74. https://doi.org/10.21831/jppfa.v9i1.38049

Yen, T. S., & Halili, S. H. (2015). Effective teaching of higher-order thinking (HOT) in education. *The Online Journal of Distance Education and E-Learning, 3*(2), 41–47. https://tojqih.net/journals/tojdel/articles/v03i02/v03i02-04.pdf