Jurnal Inovasi Teknologi Pendidikan
Volume 11, No. 2, June 2024 (172-183)
Online: http://journal.uny.ac.id/index.php/jitp

HOTS checker: Quick reviewing cognitive levels of learning outcomes using large language models

Dwi Soca Baskara, Hardika, Dio Lingga Purwodani, Nabil Muttaqin
Universitas Negeri Malang, Indonesia.
* Corresponding Author. E-mail: dwi.soca@um.ac.id

ARTICLE INFO

Article History
Received:
20 November 2023;
Revised:
15 December 2023;
Accepted:
31 December 2023;
Available online:
30 June 2024.

Keywords
Learning outcomes; Large language models; Artificial intelligence; Cognitive levels; Bloom's taxonomy; Learning design

ABSTRACT

The development of tools for efficient and effective assessment of learning outcomes is crucial in education. However, identifying the appropriate cognitive levels for learning outcomes can be challenging for educators. This study proposes to develop a tool to address this challenge by combining the strengths of large language models (LLMs) and Bloom's taxonomy. The tool can benefit educators by providing them with a streamlined reviewing process and enhancing their ability to assess learning outcomes. This research referred to prototype development models by Pressman. The research stages included communication, quick plan, modeling and quick design, construction of prototype, delivery, and feedback. The validation process involved assessing the tool's accuracy, consistency, and potential to be implemented in real educational settings by educators. The overall score obtained from the validation process is 76.92%, with the highest results coming from the categories of the tool's potentiality. It demonstrates its potential as a valuable educational tool. The insights gained from the expert validation serve as a crucial guidepost for future iterations of the tool, aligning them more closely with the goals of enhancing learning outcomes in educational settings.

How to cite:
Baskara, D.S., Hardika, Purwodani, D.L. & Muttaqin, N. (2024). HOTS checker: Quick reviewing cognitive levels of learning outcomes using large language models. *Jurnal Inovasi Teknologi Pendidikan, 11*(2), 172-183. https://doi.org/10.21831/jitp.v11i2.67174

INTRODUCTION

Learning design is a thoughtful and purposeful approach to organizing educational experiences to achieve specific learning outcomes (Koh, 2022). It involves thoughtfully considering different components, including content, assessment techniques, instructional strategies, and the integration of technology, to create an engaging and meaningful learning environment (Rosson, 2014). By employing effective learning design principles, educators can maximize students' learning activities so that they align with the desired goals and objectives of the course or program (Whitelock & Rienties, 2016). Furthermore, a well-designed learning experience can enhance students' understanding and retention of knowledge (Amien & Hidayatullah, 2023). By incorporating elements such as instructional objectives, course details, teaching plans, and learning outcomes, educators can create a clear and systematic structure for their courses.

One fundamental aspect of learning design is the establishment of clear and measurable learning outcomes (Wei et al., 2021). Learning outcomes articulate what students are expected to know, understand, or be able to do by the end of a course or instructional module (Harden, 2002). These outcomes serve as a crucial foundation for designing the entire learning experience and providing learners with a clear understanding of the goals and expectations (Albatti, 2023; Mehany & Gebken, 2021). In addition to this foundational element in the learning design theory framework includes three variables: conditions, treatments, and results (Schunk, 2012).

Conditions refer to the specific factors, such as the learning environment, resources, and support, that affect the learning experience. By understanding the conditions in which students will be learning, educators can make informed decisions about the instructional strategies and resources to employ (Bowman, 2022). Treatments involve the deliberate design and implementation of various instructional activities and interventions to facilitate learning. This includes selecting appropriate teaching methods, technologies, and assessments that align with the desired learning outcomes. Educators must carefully consider the needs and preferences of their learners when designing treatments to ensure that they are engaging and effective (Abulhul, 2021). Results refer to the expected learning outcomes and the assessment of student achievement. By clearly defining the desired results, educators can develop appropriate and meaningful assessments to measure student learning. These assessments can take various forms, such as quizzes, projects, or presentations, and should align with the intended learning outcomes (Ramanathan, 2022).

When defining learning outcomes, it is essential to identify the cognitive levels at which these outcomes operate (Maxwell, 2021). Cognitive levels represent the depth of thinking and complexity required to achieve specific learning objectives. Bloom's Taxonomy provides a framework for categorizing cognitive levels, ranging from lower-order thinking skills (e.g., remembering and understanding) to higher-order thinking skills (e.g., analyzing and creating) (Krathwohl, 2002). Research has shown that students are more likely to be motivated, satisfied, and engaged when they are presented with learning outcomes designed at higher levels of cognitive demand (Crichton & Kinsel, 2003). This underscores the significance of creating challenging and intellectually stimulating objectives.

By implementing effective learning outcomes that align with these principles, educators can create meaningful experiences for learners. Designing a student-centered environment involves clarity in course expectations and outcomes while also developing a community where interaction is encouraged. Ensuring success in following this design-based approach to enhancing teaching practices requires careful consideration of various factors such as assessment models for evaluating student achievement along with systematic evaluation techniques aimed at measuring impacts on diversity.

However, identifying the appropriate cognitive levels for learning outcomes can be challenging for educators. It requires effective measurement methods and assessment tools, including those for high-level cognitive skills, the affective domain, and the psychomotor field (Goel et al., 2021). Furthermore, determining whether an outcome should require lower-level or higher-level cognitive skills depends on various factors, including the subject matter, the level of the course, and the student's prior knowledge and abilities (Krathwohl, 2002).

This process often involves repetitive activities, such as administering the same test to different groups of students or grading numerous similar assignments (Olsen et al., 2019). Additionally, instructors must strike a balance between setting challenging goals that promote deep learning and ensuring that the objectives are achievable and attainable (Troitschanskaia et al., 2019).

Large language models (LLMs) have demonstrated remarkable capabilities in natural language processing and understanding (Pallagani et al., 2023). Large language models are Al-powered systems that can process and generate human-like text. These models have been trained on massive amounts of data to understand and generate language, making them powerful tools in various domains, including education.

Large language models (LLMs) have shown excellent text generation capabilities, capable of generating fluent human-like responses for many downstream tasks (Xiao & Shan, 2023). LLMs can mimic the human translation process and improve translation quality by extracting translation-related knowledge such as keywords, topics, and relevant demonstrations (Chen et al., 2023). These models have been trained on massive amounts of data to understand and generate language (Chang et al., 2021), making them powerful tools in various domains, including education.

In recent years, the emergence of large language models has presented new possibilities for enhancing learning design in education settings (Sarsa et al., 2022). One potential application of large language models in education is the creation of personalized learning experiences (Ma et al., 2021). With their ability to analyze and generate text, these models can adapt instruction context to meet the unique needs and preferences of individual learners (Muse et al., 2023).

By considering factors such as students' characteristics and subject matter, educators can integrate technology seamlessly into the learning process (Ramírez & Gerena, 2010). This integration serves not just as a tool for critical thinking development but also supports the practice of learning design, which involves creating, managing, and evaluating various learning activities with the aid of technology (Arcas, 2022). By leveraging large language models, educators can create tailored learning experiences that cater to the specific needs and interests of each student, maximizing their engagement and learning outcomes (Park et al., 2019).

Large language models (LLMs) work by using semantic information to process and understand text (Gilbert et al., 2023). LLMs have revolutionized various tasks such as information retrieval, question answering, summarization, and code generation (Zhang, 2021). They can effectively compress and reconstruct text while preserving the semantic essence of the original text.

However, while it operates based on semantic information, it can also be a limitation. The information or analysis results provided by LLMs may be broad and, at times, unpredictable (Tamkin et al., 2021). This unpredictability stems from the diverse sources of data the models have been trained on, including a wide array of internet text, which can encompass both accurate and inaccurate information. Consequently, when educators rely on LLMs to assist in the evaluation and categorization of cognitive levels in learning outcomes, there is the inherent risk of receiving overly generalized or even misleading guidance.

To address these limitations and enhance the precision of insights generated by LLMs, it is crucial to provide explicit context (Ratner et al., 2023). By specifying the context, it narrows down the scope of responses generated by LLMs, making them more applicable to educational environments. Contextual cues like subject matter, course level, or educational framework can help in this regard.

On the other hand, Bloom's taxonomy is widely used as a framework to classify educational objectives and assess learning outcomes (Alhazmi et al., 2015; Zorluoğlu & Güven, 2020). It provides a structured approach to understanding the complexity of cognitive processes involved in learning. Bloom's taxonomy consists of six levels, each representing a different cognitive skill: remembering, understanding, applying, analyzing, evaluating, and creating (Krathwohl, 2002).

Table 1. Six Levels of Cognitive Skills of Bloom's Taxonomy

No.	**Cognitive Level**	**Definition**	**Action Verbs Representing Intellectual Activity in Learning Outcome**
1	Knowledge	At the Foundational Level of Bloom's Taxonomy, "Knowledge" Refers to the Cognitive Skill of Recognizing and Recalling Fundamental Facts, Terms, Concepts, and Information Related to a Subject Matter. Learners at this Stage Exhibit a Capacity to Memorize and Articulate Essential Data and Terminology.	Define, Identify, Label, List, Name, Recall, Recite, State.
2	Comprehension	“Comprehension" Denotes a Cognitive Skill where Learners Demonstrate their Ability to Grasp the Meaning, Interpretation, and Significance of the Acquired Knowledge. It Involves Understanding the Content in a way that allows for Explanation, Illustration, or Summarization of the Material.	Describe, Explain, Illustrate, Infer, Paraphrase, Summarize.

No.	Cognitive Level	Definition	Action Verbs Representing Intellectual Activity in Learning Outcome
3	Application	The Cognitive Skill of "Application" Necessitates the Practical Utilization of Knowledge and Comprehension to Address Real-World Problems, Perform Tasks, or Employ Acquired Concepts in Novel Contexts. Learners are Expected to Apply their Understanding to Solve Issues or Complete Activities Effectively.	Apply, Demonstrate, Implement, Solve, Use, Execute.
4	Analysis	"Analysis" Entails a Higher Level of Cognitive Engagement in which Learners Break Down Complex Information Into its Constituent Parts, Uncovering Underlying Structures and Identifying Relationships and Patterns. This Level Requires Examining the Material Critically and Identifying Key Elements.	Analyze, Compare, Contrast, Deconstruct, Differentiate, Investigate.
5	Synthesis	At the "Synthesis" Level, Learners Exhibit The Cognitive Ability to Create Novel Insights or Ideas by Integrating and Recombining Various Elements and Concepts Into a Coherent and Original Whole. This Entails a Creative Approach to Problem-Solving and the Generation of Innovative Solutions.	Combine, Design, Develop, Formulate, Integrate, Propose.
6	Evaluation	"Evaluation" Represents the Highest Level Of Cognitive Skills in Bloom's Taxonomy, where learners make Informed Judgments or Assessments based on Predetermined Criteria and Standards. It Involves Weighing the Merits of Various Options, Often Necessitating the Justification of Choices.	Assess, Critique, Evaluate, Judge, Justify, Prioritize.

By using Bloom's taxonomy in Table 1, educators and learners can gain a deeper understanding of the depth and complexity of learning outcomes. It helps in designing effective learning experiences by aligning instructional strategies and assessment methods with specific cognitive levels (Charoensap & Saeheaw, 2022; Goštautaitė, 2019). For example, if the learning objective is to remember information, educators can design activities that focus on memorization and recall. On the other hand, if the objective is to analyze and evaluate information, educators can create tasks that require critical thinking and problem-solving. Bloom's taxonomy also serves as a guide for curriculum development and instructional planning (Hyder & Bhamani, 2016). It ensures that learning objectives progress from lower-order thinking skills to higher-order thinking skills, promoting intellectual growth and development (Sideeg, 2016). By incorporating Bloom's taxonomy into the assessment process, educators can assess not only the acquisition of knowledge but also the application, analysis, evaluation, and creation of new ideas.

Furthermore, Bloom's taxonomy encourages learners to engage in active learning (Sobral, 2021). A recent study shows that Bloom's taxonomy promotes deeper understanding by encouraging students to analyze, apply, evaluate, and create connections between course themes, texts, and concepts (Mulcare & Shwedel, 2017). It also encourages reflection by providing a framework for categorizing and assessing different levels of thinking skills (Zamir & Jan, 2023). The use of Bloom's taxonomy in textbooks and learning activities has been found to have a positive impact on critical thinking skills, creativity, and problem-solving abilities (Pujawan et al., 2022; Stevani & Tarigan, 2023).

By combining the strengths of LLMs and Bloom's taxonomy, we propose to develop a tool to answer the challenge of identifying cognitive levels of learning outcomes. This tool provides educators with a streamlined reviewing process and enhances their ability to assess learning outcomes. By leveraging the power of LLMs, educators can generate concise and contextually appropriate content to support their instructional strategies. This not only saves time but also enables educators to personalize the review experience for each learner. Additionally, incorporating Bloom's taxonomy into the assessment process ensures a comprehensive evaluation of learning outcomes. It guides educators in designing effective learning experiences that align instructional strategies and assessment methods with specific cognitive levels. These tools are also designed to promote higher-order thinking skills, critical thinking, and problem-solving, empowering learners to achieve deeper understanding and intellectual growth.

However, while it holds potential possibilities, it is important to acknowledge that this tool is currently in the development and expert validation stage. To assess its practical utility and effectiveness, empirical testing is essential. Only by conducting thorough experiments and analyzing data can we accurately assess the effectiveness of these tools. This research contributes to increasing the cognitive level using large language models.

METHOD

This research relies on prototyping models that provide a structured framework for the iterative creation of a software or technological solution (Pressman & Maxim, 2020). In the context of this study, prototyping models have been systematically refined to address specific requirements garnered through interviews with educators depicted in Figure 1.

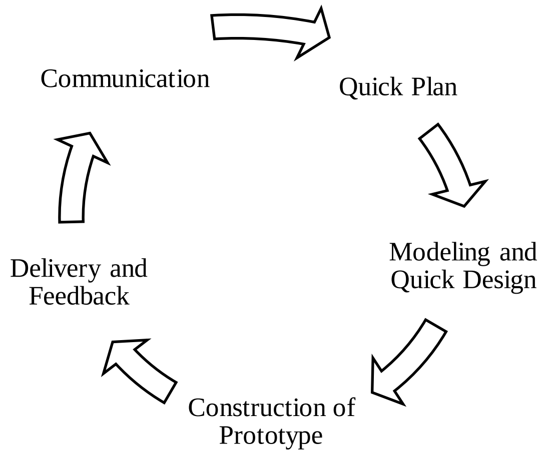

Figure 1. The Diagram of Prototype Models

1. Communication
   The first stage of the prototyping process involves customizing the features of the tool to meet educators' specific needs and preferences. Educators have emphasized the need for a tool that can effectively assess learning outcomes at an early stage in the learning plan. They also stress the importance of aligning this assessment with Bloom's Taxonomy, which categorizes cognitive levels for comprehensive evaluation. It is crucial for educators to seamlessly integrate this tool into their Learning Management Systems to ensure easy access and use. Additionally, they would like valuable feedback from the tool that addresses different cognitive levels of learning outcomes and provides suggestions to improve higher-order thinking skills. These insights will guide subsequent steps in developing an authentic student assessment system that meets all requirements.
2. Quick Plan
   A comprehensive understanding of educators' needs drives the creation of a well-defined project plan. This plan outlines specific timeframes for each development phase and incorporates key elements identified through teacher feedback. It ensures that the assessment tool aligns with

Bloom's Taxonomy, enabling educators to meet their teaching goals while fostering critical thinking and providing valuable feedback. Additionally, the plan emphasizes smooth integration within existing Learning Management Systems, simplifying access and usability for students. Resource allocation is carefully managed to ensure timely completion of the prototype while enhancing student engagement in applying knowledge and skills relevant to real-world scenarios.

3. Modeling and Quick Design
   Based on the information gathered, the design phase will be conducted to create a user-friendly and efficient interface. The wireframing process will take into account different levels of cognitive thinking according to Bloom's Taxonomy to align with educators' pedagogical goals. Additionally, careful planning will involve integrating LLMs APIs and prioritizing accurate feedback that is relevant to the context. The objective is to develop a user experience that supports educators in simplifying assessment processes while fostering advanced critical thinking skills.
4. Construction of Prototype
   During the development phase, the prototype is carefully constructed according to design and functional specifications. The coding process aims to incorporate LMS capabilities and align assessment features with learning outcomes based on Bloom's Taxonomy. This ensures a thorough analysis that supports curriculum development. Extensive testing is conducted to identify and address any technical issues, resulting in a resilient and user-friendly prototype. Collaboration with educators through feedback loops plays a crucial role in refining the tool, optimizing its effectiveness, and considering individual variations in learning styles while aligning it with their educational objectives.
5. Deployment, Delivery, and Feedback
   To ensure the effectiveness and usability of the prototype, it is deployed to a selected group of educators and stakeholders. Their valuable feedback on factors such as ease of use, efficiency, and impact on learning outcomes is carefully gathered. In addition, expert validation sessions involving educators are conducted to further refine the tool's assessment capabilities for cognitive levels and promote higher-order thinking skills. The received feedback and validation results serve as important inputs for continuous improvements, ensuring that the Quick Reviewing Learning Outcomes using the LLMs tool remains aligned with evolving needs and goals in education.

RESULTS AND DISCUSSION

Results

The need assessment is conducted by interviewing some teachers. Table 2 shows the results of the interview.

Table 2. Need Assessment Results

No.	Question	Answer
1	How do Educators Approach Learning Plan Development?	Educators Prioritize Addressing Learning Outcomes and Dedicate Considerable Time to Assessing Them at an Early Stage.
2	What Emphasis do Educators Place on Assessment Alignment?	Educators Stress the Importance of Aligning Assessments With Bloom's Taxonomy to Ensure Comprehensive Evaluation of Cognitive Levels.
3	How do Educators Aim to Integrate Assessment Tools Into their Systems?	Educators Find It Crucial to Seamlessly Integrate Assessment Tools Into their Learning Management Systems for Easy Access and Usability.
4	What Feedback Do Educators Seek from Assessment Tools?	Educators Seek Valuable Feedback That Addresses Different Cognitive Levels of Learning Outcomes and Provides Suggestions to Enhance Higher-Order Thinking Skills.

After conducting a need assessment with educators, we have decided to develop a web browser extension that can seamlessly integrate into an LMS. The primary function of this tool will be to analyze and assess the cognitive level of learning outcomes. Additionally, it will provide valuable feedback in the form of recommendations aimed at fostering high-order thinking skills.

To begin development, a quick plan has been outlined. In the initial stage, determine the flow diagram in Figure 2 to visualize a sequence of actions within the tools. Following that, the appropriate LLMs models to be used are determined and the necessary backend infrastructure will be developed in stage two. Stage three involves designing an intuitive user interface that can be easily integrated as a web browser extension. Finally, stage four will focus on integrating this tool directly into existing LMS platforms.

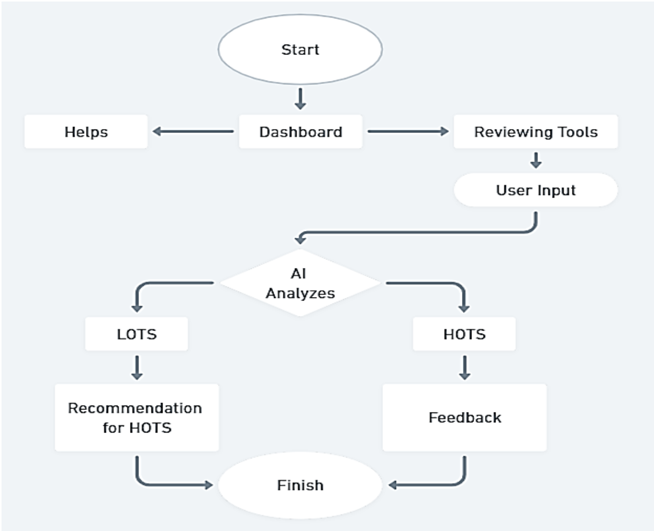

Figure 2. The Diagram of the HOTS Checker

To begin, users should access the installed tools. Then, they can navigate to the review menu or help section. Within the reviewing menu, users can input their learning outcomes and click on the send button. The system will then analyze the input and provide a response that includes the cognitive levels of the learning goals. Additionally, it will offer reasoning for its judgment. If the cognitive levels indicate lower-order thinking skills, recommendations will be generated to enhance higher-order thinking skills. A visual preview of these tools is displayed in Figure 3.

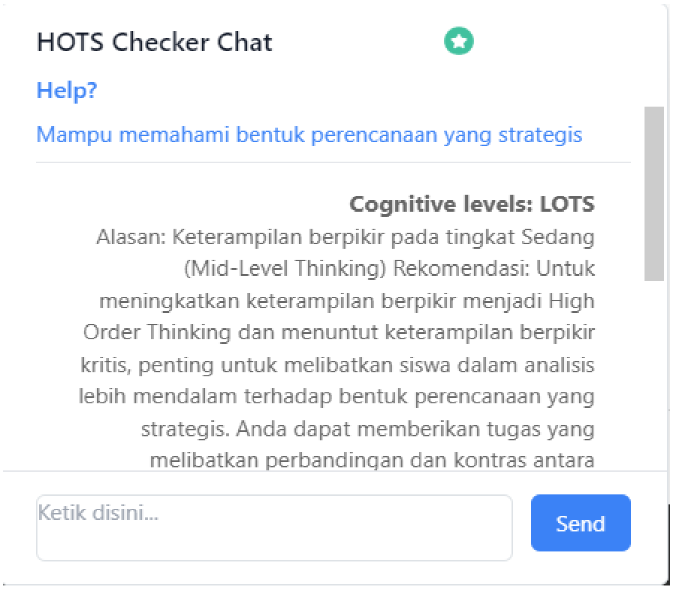

Figure 3. The Interface of HOTS Checker

After developing the tools, the next step in the process is expert validation. This serves as the final step in prototyping models. Expert validation plays a pivotal role in assessing the effectiveness and viability of the developed models. By involving the experts, we can gather valuable insights, identify potential flaws or improvements, and ensure that the prototype aligns with the desired objectives. We can refine and validate the model before moving forward with its implementation.

The validation of the tool was conducted by a learning assessment expert. The validation process involved assessing the tool's accuracy, consistency, and potential to be implemented in real educational settings by educators. The expert provided feedback using a questionnaire format, with each item rated on a scale of 1 to 4. The following results are presented in Table 3.

Table 3. Expert Validation Results

No.	Validation Category	Evaluation Aspect	Results
1	Tool's Accuracy	Accuracy of Learning Outcome Assessment	3
		Clarity of Explanations	3
		Clarity of Recommendations	4
		Alignment with Bloom's Taxonomy	3
2	Tool's Consistency	Consistency in Classification	2
		Consistency in Explanations	2
		Consistency in Recommendations	2
3	Tool's Potentiality	Usefulness for Identifying Areas of Improvement	4
		Enhancing Critical Thinking and Learning Skills	3
		Support for Better Lesson Planning	4
		Direct Application for Learning Improvement	2
		Information for Learning Improvement	4
		Adaptability to Various Learning Scenarios	4

Table 4 provides a quantitative summary of the expert validation results for each part and the overall assessment of the tool's effectiveness.

Table 4. Quantitative Summary of Expert Validation

No.	Part of Validation	Scores	Percentage (%)
1	Tool's Accuracy	13	81.25%
2	Tool's Consistency	6	50%
3	Tool's Potentiality	21	87.5%
Overall (All Parts)		**40**	**76.92%**

Discussion

The expert validation results indicate that the tool for Quick Reviewing Learning Outcomes using LLMs holds promise as a valuable educational resource. Its strengths lie in its accuracy in aligning assessments with actual skill levels and in its potential to offer valuable insights to educators for lesson planning and identifying areas of improvement in student learning skills (Caines et al., 2023). However, there are clear areas for improvement, particularly in terms of ensuring consistency in assessments, explanations, and recommendations (Raj et al., 2022). Additionally, efforts should be directed toward enhancing the tool's direct applicability for learning improvement, which received a lower rating from the expert.

Overall, these validation results provide a valuable foundation for the ongoing development and refinement of the tool. By addressing the identified areas of improvement, the tool can be further optimized to meet the needs of educators and contribute to more effective and data-informed educational practices (Sahu et al., 2022). The insights gained from expert validation serve as a crucial guidepost for future iterations of the tool, aligning it more closely with the goals of enhancing learning outcomes in educational settings.

The expert validation results offer valuable insights and lead to several key findings and recommendations. The incorporation of learning level matrices in assessments has great promise for educators (Elkins et al., 2023). They can benefit from the efficient review process and personalized experiences provided by these tools. Additionally, the ability of the tool to align assessments with skill levels represents an encouraging advancement in educational technology that is aligned with Bloom's taxonomy principles, providing a comprehensive framework for assessing learning outcomes (Caines et al., 2023).

However, the validation results also underscore the need for improvement in terms of consistency. Achieving uniformity in assessments, explanations, and recommendations is pivotal to enhancing the tool's reliability and usability (Chen et al., 2023). Addressing this aspect should be a priority in further tool development. Furthermore, the lower rating for the tool's direct applicability for learning improvement highlights an area that requires attention. Enhancements should focus on making assessment results more actionable for educators, thereby facilitating immediate improvements in the learning process.

CONCLUSION

It is crucial to acknowledge that these tools are currently in the stage of development and expert validation. Their practical usefulness and effectiveness in actual educational settings can only be determined through empirical testing and ongoing improvements. The process of development and validation is ongoing, emphasizing the importance of addressing recognized limitations and improving overall functionality. Furthermore, it is important to note that while LLMs can be a valuable tool for quick reviewing, they struggle with long-term planning and finding optimal solutions. Therefore they should not replace educators for comprehensive studying and deep learning. They are best utilized as a supplementary resource to complement traditional learning methods.

In summary, the integration of LLMs and Bloom's taxonomy into assessing and reviewing learning outcomes presents a promising new approach to education. These tools have the potential to empower educators to personalize learning experiences, align instructional strategies with cognitive levels, and ultimately improve student learning outcomes. However, it is important to recognize that these tools are still developing and require ongoing refinement and validation to fulfill their transformative potential in education. By incorporating feedback from experts and actively working on areas that need improvement, we can develop more efficient educational approaches that are based on data and prioritize the needs of learners. This will be beneficial for both educators and students alike.

REFERENCES

Abulhul, Z. (2021). Teaching strategies for enhancing student learning. *Journal of Practical Studies in Education*, *2*(3), 1–4. https://doi.org/10.46809/jpse.v2i3.22

Albatti, H. (2023). A review of intended learning outcomes of English lessons and learning motivation. *AWEJ: Arab World English Journal*, *14*(2), 205–220. https://doi.org/10.24093/awej/vol14no2.15

Alhazmi, A. K., Zafar, H., & Al-Hammadi, F. (2015). Framework for integrating outcome-based assessment in online assessment: Research in progress. *2015 Science and Information Conference (SAI)*, 217–221. https://doi.org/10.1109/SAI.2015.7237147

Amien, M. S., & Hidayatullah, A. (2023). Assessing students' metacognitive strategies in e-learning and their role in academic performance. *Jurnal Inovasi Teknologi Pendidikan*, *10*(2), 158–166. https://doi.org/10.21831/jitp.v10i2.60949

Arcas, B. A. y. (2022). Do large language models understand us? *Daedalus*, *151*(2), 183–197. https://www.jstor.org/stable/48662035

Bowman, R. F. (2022). Cornerstones of productive teaching and learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *95*(2), 57–63. https://doi.org/10.1080/00098655.2022.2033671

Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, O., Yuan, Z., Elliott, M., Moore, R., Bryant, C., Rei, M., Yannakoudakis, H., Mullooly, A., Nicholls, D., & Buttery, P. (2023). On the application of large language models for language teaching and assessment technology. *AIED2023 Workshop: Empowering Education with LLMs - the Next-Gen Interface and Content Generation*, 1–25. https://doi.org/10.48550/arXiv.2307.08393

Chang, E., Demberg, V., & Marin, A. (2021). Jointly improving language understanding and generation with quality-weighted weak supervision of automatic labeling. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 818–829. https://doi.org/10.18653/v1/2021.eacl-main.69

Charoensap, K., & Saeheaw, T. (2022). Customer experiences identification process using Bloom taxonomy and customer knowledge management. *2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, 122–127. https://doi.org/10.1109/ECTIDAMTNCON53731.2022.9720405

Chen, S., Gao, S., & He, J. (2023). Evaluating factual consistency of summaries with large language models. *Computation and Language*, 1–12. https://doi.org/10.48550/arXiv.2305.14069

Crichton, S., & Kinsel, E. (2003). Learning plans as support for the development of learner identity: A case study in rural Western Canada. *Journal of Adult and Continuing Education*, *8*(2), 213–226. https://doi.org/10.7227/JACE.8.2.7

Elkins, S., Kochmar, E., Cheung, J. C. K., & Serban, I. (2023). How useful are educational questions generated by large language models? *AIED Late Breaking Results 2023*, 536–542. https://doi.org/10.48550/arXiv.2304.06638

Gilbert, H., Sandborn, M., Schmidt, D. C., Spencer-Smith, J., & White, J. (2023). Semantic compression with large language models. *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 1–8. https://doi.org/10.1109/SNAMS60348.2023.10375400

Goel, N., Deshmukh, K., Patel, B. C., & Chacko, S. (2021). Tools and rubrics for assessment of learning outcomes. In *Assessment Tools for Mapping Learning Outcomes With Learning Objectives* (pp. 1–44). IGI Global. https://doi.org/0.4018/978-1-7998-4784-7.ch013

Goštautaitė, D. (2019). Principal component analysis and Bloom taxonomy to personalise learning. *Proceedings of EDULEARN19 Conference*, 2910–2920. https://doi.org/10.21125/EDULEARN.2019.0780

Harden, R. M. (2002). Learning outcomes and instructional objectives: is there a difference? *Med Teach*, *24*(2), 151–155. https://doi.org/10.1080/0142159022020687

Hyder, I., & Bhamani, S. (2016). Bloom's taxonomy (cognitive domain) in higher education settings: Reflection brief. *Journal of Education and Educational Development*, *3*(2), 288–300. https://jmsnew.iobmresearch.com/index.php/joeed/article/view/198

Koh, J. H. L. (2022). Designing for designerly ways of knowing: Creating learning design futures in higher education. In *Design Praxiology and Phenomenology*. Springer. https://doi.org/10.1007/978-981-19-2806-2_11

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

Ma, Z., Dou, Z., Zhu, Y., Zhong, H., & Wen, J.-R. (2021). One chatbot per person: Creating personalized chatbots based on implicit user profiles. *SIGIR*, 1–10. https://doi.org/10.1145/3404835.3462828

Maxwell, G. S. (2021). Defining and assessing desired learning outcomes. In *The Enabling Power of Assessment* (pp. 1–399). Springer. https://doi.org/10.1007/978-3-030-63539-8_3

Mehany, M. S. H. M., & Gebken, R. (2021). Assessing the importance and cognition level of ACCE's student learning outcomes: Industry, educator, and student perceptions. *International Journal of Construction Education and Research*, *17*(4), 333–351. https://doi.org/10.1080/15578771.2020.1777487

Mulcare, D. M., & Shwedel, A. (2017). Transforming Bloom's taxonomy into classroom practice: A practical yet comprehensive approach to promote critical reading and student participation. *Journal of Political Science Education*, *13*(2), 121–137. https://doi.org/10.1080/15512169.2016.1211017

Muse, H., Bulathwela, S., & Yilmaz, E. (2023). Pre-training with scientific text improves educational question generation (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 16288–16289. https://doi.org/10.1609/aaai.v37i13.27004

Olsen, D. R., Keith, J., & Rosenberg, J. (2019). The quest for increased learning: Systematically aligning and assessing learning outcomes. *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 314–319. https://doi.org/10.1109/IIAI-AAI.2019.00070

Pallagani, V., Muppasani, B., Murugesan, K., Rossi, F., Srivastava, B., Horesh, L., Fabiano, F., & Loreggia, A. (2023). Understanding the capabilities of large language models for automated planning. *Computer Science*, 1–12. https://doi.org/10.48550/arXiv.2305.16151

Park, H. W., Grover, I., Spaulding, S., Gomez, L., & Breazeal, C. (2019). A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education. *Proceedings of the AAAI Conference on Artificial Intelligence*, 687–694. https://doi.org/10.1609/aaai.v33i01.3301687

Pressman, R., & Maxim, B. (2020). *Software engineering: A practitioner's approach*. McGraw Hill.

Pujawan, I. G. N., Rediani, N. N., Antara, I. G. W. S., Putri, N. N. C. A., & Bayu, G. W. (2022). Revised Bloom taxonomy-oriented learning activities to develop scientific literacy and creative thinking skills. *Jurnal Pendidikan IPA Indonesia*, *11*(1), 47–60. https://doi.org/10.15294/jpii.v11i1.34628

Raj, H., Rosati, D., & Majumdar, S. (2022). Measuring reliability of large language models through semantic consistency. *NeurIPS 2022 ML Safety Workshop*, 1–7. https://doi.org/10.48550/arXiv.2211.05853

Ramanathan, C. (2022). Technologies for teaching learning process, its evaluations and assessments. In *Development of Employability Skills Through Pragmatic Assessment of Student Learning Outcomes* (pp. 1–18). IGI Global. https://doi.org/10.4018/978-1-6684-4210-4.ch003

Ramírez, M. D. V., & Gerena, L. (2010). Bilingual education from learner perspectives. In *Handbook of Research on Bilingual and Intercultural Education* (pp. 1–19). IGI Global. https://doi.org/10.4018/978-1-7998-2588-3.ch017

Ratner, N., Levine, Y., Belinkov, Y., Ram, O., Magar, I., Abend, O., Karpas, E., Shashua, A., Leyton-Brown, K., & Shoham, Y. (2023). Parallel context windows for large language models. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 6383–6402. https://doi.org/10.18653/v1/2023.acl-long.352

Rosson, M. B. (2014). Learning by design. In *Innovative Practices in Teaching Information Sciences and Technology* (pp. 75–83). Springer. https://doi.org/10.1007/978-3-319-03656-4_8

Sahu, P., Cogswell, M., Gong, Y., & Divakaran, A. (2022). Unpacking large language models with conceptual consistency. *ICLR*, 1–14. https://doi.org/10.48550/arXiv.2209.15093

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. *ICER '22: Proceedings of the 2022 ACM Conference on International Computing Education Research*, 27–43. https://doi.org/10.1145/3501385.3543957

Schunk, D. H. (2012). *Learning theories: An educational perspective*. Pearson.

Sideeg, A. (2016). Bloom's taxonomy, backward design, and Vygotsky's zone of proximal development in crafting learning outcomes. *International Journal of Linguistics*, *8*(2), 158–186. https://doi.org/10.5296/ijl.v8i2.9252

Sobral, S. R. (2021). Bloom's taxonomy to improve teaching-learning in introduction to programming. *International Journal of Information and Education Technology*, *11*(3), 148–153. https://doi.org/10.18178/ijiet.2021.11.3.1504

Stevani, M., & Tarigan, K. E. (2023). Evaluating English textbooks by using Bloom's taxonomy to analyze reading comprehension question. *SALEE: Study of Applied Linguistics and English Education*, *4*(1), 1–18. https://doi.org/10.35961/salee.v0i0.526

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the capabilities, limitations, and societal impact of large language models. *Computation and Language*, 1–8. https://doi.org/10.48550/arXiv.2102.02503

Troitschanskaia, O. Z., Schlax, J., Jitomirski, J., Happ, R., Thees, C. K., Brückner, S., & Pant, H. A. (2019). Ethics and fairness in assessing learning outcomes in higher education. *Higher Education Policy*, *32*(1), 537–556. https://doi.org/10.1057/s41307-019-00149-x

Wei, X., Saab, N., & Admiraal, W. (2021). Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review. *Computers & Education*, *163*(1), 1–24. https://doi.org/10.1016/j.compedu.2020.104097

Whitelock, D., & Rienties, B. (2016). #Design4learning: Designing for the future of higher education. *Journal of Interactive Media in Education*, *2016*(1), 1–3. https://doi.org/10.5334/JIME.417

Xiao, L., & Shan, X. (2023). PatternGPT: A pattern-driven framework for large language model text generation. *Computation and Language*, 1–14. https://doi.org/10.48550/arXiv.2307.00470

Zamir, S., & Jan, H. (2023). Assessment of papers of English of Sukkur BISE Sindh, Pakistan: An exploration of the reflection of Bloom's taxonomy. *Voyage Journal of Education Studies*, *3*(1), 220–240. https://doi.org/10.58622/vjes.v3i1.41

Zhang, H. (2021). Transfer training from smaller language model a preprint. *ArXiv*, 1–7.

Zorluoğlu, S. L., & Güven, Ç. (2020). Analysis of 5th grade science learning outcomes and exam questions according to revised Bloom taxonomy. *Journal of Educational Issues*, *6*(1), 58–69. https://doi.org/10.5296/jei.v6i1.16197