Analysis of the Impact of ChatGPT Usage in Direct Current Physics Material on Enhancing Students' Higher-Order Thinking Skills (HOTS) during the Merdeka Belajar

Zakaria Al Farizi1*, Sahlan1, Erwin Prasetyo1, Nadya Fitriani2, Elyakim Mau Lalang2, Hamzarudin Hikma Tiar1, and Zaina Anwar1

1Physics Education, Universitas Muhammadiyah Maumere, Waioti, Indonesia
2MAN Lembata, Desa Kelikur, Indonesia

*Zakariaalfarizi797@gmail.com

DOI:10.20527/bipf.v12i1.17603

Received: 5 October 2023  Accepted: 5 January 2024  Published: 1 February 2024

Abstract

This study examines the influence of using ChatGPT to enhance students' higher-order thinking Skills (HOTS), particularly within the context of the "Merdeka Belajar" (Freedom to Learn) era in Indonesia. This inquiry arises from Indonesian students' low literacy and numeracy skills, as indicated by the PISA assessment. Employing a quasi-experimental method, the research analyzes the changes in students' HOTS before and after the ChatGPT-assisted instruction intervention within the "Merdeka Belajar" curriculum framework. The study's findings reveal a significant improvement in students' HOTS within the experimental group that utilized ChatGPT, as compared to the control group. Data analysis underscores the positive impact of integrating ChatGPT into "Merdeka Belajar"-based teaching methods, substantially enhancing students' HOTS. These results align with prior literature highlighting the role of AI technology in advancing education and fostering higher-order thinking abilities. It is imperative to consider that the effective utilization of AI technology should be complemented by appropriate teaching approaches to prepare students for future challenges. The implications of this research indicate that ChatGPT can be considered an innovative strategy to support Merdeka Belajar-based learning in Indonesia.

Keywords: AI technology; ChatGPT; Higher Order Thinking Skills (HOTS); Independent learning; Student ability

© 2024 Berkala Ilmiah Pendidikan Fisika


INTRODUCTION

Education plays a crucial role in shaping a nation's future. Quality education imparts knowledge and molds the character and skills required to face future challenges (Idris et al., 2012). However, data from the Programme for International Student Assessment (PISA) in 2019 indicates that Indonesia faces significant challenges in education, particularly in mathematics and literacy. Indonesia's ranking, which is sixth from the bottom among 79 countries, highlights the need for
significant improvements. The Minister of Education and Culture has launched the Independent Learning program as a breakthrough to enhance the quality of education in Indonesia (Mustaghfiroh, 2020; Suharno et al., 2023). The Merdeka Belajar program emphasizes the development of students’ creativity, skills, and mastery of literacy and numeracy. Within this framework, the Independent Learning Curriculum allows students and schools to design learning experiences that consider students’ sensitivity to social phenomena, enabling them to understand the learning process they need to undertake (Ramadani & Desyandari, 2022).

Society faces increasingly complex global challenges, and students must develop skills, competencies, and abilities as their initial capital before entering society (Murtono & Shufa, 2022). To meet this demand, the Higher Order Thinking Skills (HOTS) method can be employed (Sofyan, 2019). HOTS is a cognitive process that elevates students’ thinking abilities, including critical thinking when receiving information, creative thinking in problem-solving, and making the best decisions in complex situations (Dinni, 2018). Zajuli’s research (2019) indicates that HOTS is one of the skills needed to address current global challenges.

Anderson & Krathwohl (2001) revealed that three cognitive levels measure HOTS within the revised Bloom’s taxonomy, namely C4 (Ability to analyze), C5 (Ability to evaluate), and C6 (Ability to create) (Saraswati & Agustika, 2020; Yusuf & Widyaningsih, 2018). HOTS components generally consist of four elements, as depicted in the following diagram in Figure 1.

Figure 1 Higher Order Thinking Skills (HOTS) capability (Serevina et al., 2019)

Figure 1 summarizes various aspects of thinking skills, including analysis, argumentation, metacognition, and critical problem-solving abilities. Various efforts have been made to develop these four essential components of HOTS skills. Anisa (2020) utilized virtual simulations in her research, Sari (2022) developed PBL-based E-student worksheet, and Sani (2020) applied the 5E learning cycle model. However, despite various research endeavors, artificial intelligence (AI) technology has not been extensively explored.

Recent technological advances have ushered in the era of artificial intelligence, known as AI (Yeadon et al., 2023). AI has found applications in various sectors, offering significant benefits in meeting human needs (Susdarwono, 2021). A prominent example of this development is advanced language models like the Generative Pre-trained Transformer (GPT) (Faiz & Kurniawaty, 2023; Haleem et al., 2022). ChatGPT is a real-world example, extensively trained to process text on a large scale (Iyengar et al., 2023; Kasneci et al., 2023). ChatGPT has attracted public attention, including in the educational realm (Memarian & Doleck, 2023). While ChatGPT holds the potential to advance the academic world, it also raises various concerns (Greitemeyer & Kastenmüller, 2023; Lund & Wang, 2023).

Education is generally recognized as one of the sectors most impacted by
ChatGPT, creating a significant transformational impact (Dwivedi et al., 2023). Research on using ChatGPT in educational contexts has been extensive, revealing various facts encompassing both positive and negative impacts (Dahlkemper et al., 2023a; Kortemeyer, 2023). According to Lund & Wang (2023), ChatGPT represents a form of artificial intelligence that has the potential to contribute to the development of the education sector but carries several concerns. Research by Setiawan & Luthfiyani (2023) also highlights ChatGPT's substantial potential in advancing education. However, it is essential to remember that ChatGPT indirectly changes how humans interact with machines, opening up new opportunities in this sector's development (Ausat et al., 2023). This perspective aligns with Supriyadi's (2022) views, but Supriyadi emphasizes the importance of reinforcing academic moral values to avoid excessive dependence.

Specifically, there has been no prior research that has specifically explored ChatGPT's potential to enhance HOTS. Yet, ChatGPT has significant potential to serve as a support system for students in solving HOTS problems. One of its key abilities is text analysis. ChatGPT can summarize information from various subjects and generate new ideas based on the input provided. In this regard, ChatGPT can process text precisely and even analyze HOTS problems. Some experts have suggested that ChatGPT can create high-quality HOTS problems (Aziza et al., 2023). Furthermore, based on the research conducted by Gembong Baskoro et al. (2023) it can be revealed that with the assistance of AI, the cognitive domain of HOTS can be significantly enhanced. Therefore, ChatGPT has the potential to translate HOTS problems and assist students in overcoming the challenges they face. Additionally, ChatGPT also has the potential to provide broader assistance in solving HOTS problems. This includes analyzing problems, identifying patterns and relationships at the C4 level, evaluating various solutions or arguments at the C5 level, and providing inspiration and creative ideas for designing new solutions or projects at the C6 level. With these various capabilities, ChatGPT can be a valuable partner in helping students better understand, analyze, and solve HOTS problems while developing their higher-order thinking skills.

The background of the problem driving this research has led to an in-depth analysis of the impact of using ChatGPT in enhancing students' HOTS, particularly in the context of the "Merdeka Belajar" era. This study presents significant insights by harnessing the analytical potential of ChatGPT to assess its influence in developing students' HOTS. The novelty brought by this research not only summarizes the advancements in technology and education but also opens new opportunities in understanding and implementing artificial intelligence for more meaningful education. In the spirit of "Merdeka Belajar," ChatGPT as a learning aid holds promise in creating a generation of students better equipped to tackle high-level thinking challenges and play a vital role in advancing education.

METHOD
The quasi-experimental method is the most appropriate method for observing the impact of using ChatGPT on students' HOTS abilities during the Merdeka Belajar era. Quasi-experimental research methods are employed, which involve controlling at least one variable and observing its effect on the dependent variable (Yuwanto, 2019). Gay (1981) argued that experimental research is the most appropriate approach for examining cause-and-effect relationships between variables. This research adopts a
quantitative approach, collecting primary data directly from the study.

To ensure valid results, a sample of two test classes was randomly selected using a pretest-posttest research design (Sugiyono, 2019). The design of the research referred to is outlined in Table 1.

<table>
<thead>
<tr>
<th>Class</th>
<th>Pre-test Treatment</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eksperimen</td>
<td>X1</td>
<td>Y1</td>
</tr>
<tr>
<td>Control</td>
<td>X2</td>
<td>Y2</td>
</tr>
</tbody>
</table>

Explanation:
X1 = Eksperimen group pre-test
X2 = Control group pre-test
X = Treatment
Y1 = Eksperimen group post-test
Y2 = Control group post-test

Data related to students' HOTS abilities before and after using ChatGPT were analyzed to assess significant improvements. The research was conducted with class XII Science 1 and Science 2 students at MAN Lembata who had implemented the Merdeka Belajar curriculum. The experimental group received instruction using the Project-Based Learning (PjBL) method with the assistance of ChatGPT as a medium. Meanwhile, the control group received instruction using the PjBL method without ChatGPT. The focus of this research was on the physics of direct current electricity.

The research lasted one month, from July 20 to August 20, 2023. The saturated sampling technique, involving all population members, was employed. The research instrument comprised ten essay questions aligned with Bloom's taxonomy levels C4 to C6. Before data analysis, these questions were subjected to validity, reliability, discriminating power, and difficulty testing.

An effect size analysis was conducted to achieve the research objectives, which included identifying the impact of the treatment provided and comparing the two sample groups (control and experimental). This effect size analysis gives insights into the extent of influence one variable has on another, indicating the magnitude of improvement that occurs. Consequently, this analysis offers a greater insight into the significance and essence of using ChatGPT in enhancing students' HOTS abilities.

RESULT AND DISCUSSION

Pre-test and Postest
The experimental and control classes were administered HOTS questions at the outcome of the learning process as a pre-test to assess students' initial abilities. Subsequently, a post-test was administered to both classes to evaluate whether there were any changes in students' abilities. The scores obtained for both classes are presented in Table 2.

Table 2 The description of the score acquisition for both test classes.

<table>
<thead>
<tr>
<th>Data</th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
</tr>
<tr>
<td>Highest Value</td>
<td>36</td>
<td>90</td>
</tr>
<tr>
<td>Lowest Value</td>
<td>10</td>
<td>61</td>
</tr>
<tr>
<td>Mean</td>
<td>23.7</td>
<td>76.8</td>
</tr>
<tr>
<td>Stdv</td>
<td>7.4</td>
<td>8.4</td>
</tr>
</tbody>
</table>

Table 2 describes the scores acquired from the two test groups during the pre-test and post-test stages. These values offer an initial perspective on students' abilities before and after the treatment. During the pre-test stage, the experimental group achieved a score ranging from a high of 36 to a low of 10, while the control group had a high score of 38 and a low score of 10. After undergoing the learning treatment, which included the application of ChatGPT and the Independent Learning concept, there was an increase in scores during the post-test stage.

Upon conducting a more detailed analysis, a comparison of the initial average scores (pre-test) of the two groups revealed an insignificant difference, averaging around 23.7. However, a more substantial difference
emerged in the average scores after the treatment (post-test). The experimental group achieved an average score of 76.8, while the control group averaged 70.45. This indicates a more significant improvement in HOTS abilities in the experimental group following the treatment.

To gain a comprehensive understanding, data on the frequency distribution of pre-test and post-test scores in both groups were also analyzed, and the results are presented in Table 3.

Table 3 The frequency data for categorizing the scores of both test classes

<table>
<thead>
<tr>
<th>No</th>
<th>Data</th>
<th>Eksperimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre-test</td>
</tr>
<tr>
<td>1</td>
<td>Very High</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Very Low</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 3 displays the outcomes of the frequency categorization of values in the two test groups before and after the treatment. Grades such as very high, high, fair, low, and very low were assigned to each group. Previously to the treatment (pre-test), both groups had 20 students classified in the Very Low category. However, a notable change occurred after the implementation of ChatGPT in the experimental group. Five students from the experimental group achieved the Very High category, whereas the control group had only one student. These findings suggest that using ChatGPT within the context of Merdeka Belajar can enhance students' HOTS abilities, particularly in the higher-level categories, while reducing the number of students with very low abilities.

Normality and Homogeneity Tests

Normality and homogeneity tests were conducted to evaluate sample data distribution and the variance equality among test groups. The objective was to determine whether the data conforms to a normal distribution. The data normality test was executed using the Shapiro-Wilk one-sample test via SPSS software, with a significance level of 5%. Comprehensive information regarding the procedures and results of the normality test can be found in Tables 4 and 5.

Table 4 The criteria for normality test

<table>
<thead>
<tr>
<th>Sig</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig &gt; 0.05</td>
<td>Normally Distributed</td>
</tr>
<tr>
<td>Sig &lt; 0.05</td>
<td>Not Normally distributed</td>
</tr>
</tbody>
</table>

Table 5 Results of Normality Test

<table>
<thead>
<tr>
<th>Sample Group</th>
<th>Sig</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test Eksperimental</td>
<td>0.094</td>
<td>Normal</td>
</tr>
<tr>
<td>Post-test Eksperimental</td>
<td>0.540</td>
<td>Normal</td>
</tr>
<tr>
<td>Pre-test Control</td>
<td>0.293</td>
<td>Normal</td>
</tr>
<tr>
<td>Post-test Control</td>
<td>0.148</td>
<td>Normal</td>
</tr>
</tbody>
</table>

Table 4 displays the outcomes of the normality test for each sample group, which includes the pre-test and post-test for both the experimental and control groups. Based on the test results, it can be concluded that all sample groups have a significance value (Sig) greater than 0.05. This indicates that the data in this study can be considered to follow a normal distribution. This aligns with the criteria for the normality test employed, where a value above 0.05 signifies that the data approximates a normal distribution.

In detail, the findings of the normality test for the pre-test and post-test in both the experimental and control groups yielded Sig values greater than 0.05. Therefore, all the data utilized in this research can be categorized as having a
normal distribution. These findings instill confidence that the statistical analyses conducted on the data will yield more reliable and valid results.

Subsequently, a variance homogeneity test was conducted on the normally distributed data, following the method described by Erpina et al. (year). This test aims to verify whether the variations between groups in pre-test and post-test scores "originate from similar variances. Specifically, it examines whether there is a significant difference in the variation between the pre-test and post-test scores of the two sample groups. The variance homogeneity test was executed using the homogeneity of variances method in SPSS software, with a significance level of 5%. The results from the variance homogeneity test will offer further insights into the similarities or differences in variation between the pre-test and post-test scores in the two sample groups. Table 5 is presented as a summary of the data homogeneity test for the final results.

<table>
<thead>
<tr>
<th>Table 6 Criteria for normality test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig</td>
</tr>
<tr>
<td>Sig &gt; 0.05</td>
</tr>
<tr>
<td>Sig &lt; 0.05</td>
</tr>
</tbody>
</table>

If the Sig value obtained is greater than 0.05, it can be concluded that the data exhibits homogeneity. Conversely, if the Sig value is less than 0.05, the data is considered not homogeneous. The results of the data homogeneity test can be found in Table 7 as follows.

<table>
<thead>
<tr>
<th>Tabel 7 Homogeneity test result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Group</td>
</tr>
<tr>
<td>Pre-test Eksperimental and Control</td>
</tr>
<tr>
<td>Post-test Eksperimental and Control</td>
</tr>
</tbody>
</table>

Based on the guidelines in Table 5, which dictate how to interpret the results of the homogeneity test, it can be concluded that the data from both groups in the "pre-test and post-test" exhibit homogeneity.

**Hypothesis Testing**

Hypothesis testing is the process of examining hypotheses that are formulated before conducting research. In this research, the summarized hypotheses are as follows:

- **Ho** = There is no significant difference in enhancing students' HOTS skills between the experimental and control groups.

- **Ha** = significant difference in enhancing students' HOTS skills between the experimental and control groups “.

Hypothesis testing is conducted using an independent sample t-test”, considering this research involves two test groups. However, several prerequisites must be met before conducting this test, such as normal data distribution and variance homogeneity. A significance level of 5% is used, and the criteria for interpreting the test results can be found in Table 8.

<table>
<thead>
<tr>
<th>Table 8 Hypothesis testing criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig</td>
</tr>
<tr>
<td>Sig &gt; 0.05</td>
</tr>
<tr>
<td>Sig &lt; 0.05</td>
</tr>
</tbody>
</table>

There is no significant difference in the improvement of students' HOTS
abilities between the experimental and control groups.

Sig < 0.05  
Ho rejected, Ha accepted

There is a significant difference in the improvement of students' HOTS abilities between the experimental and control groups. Table 8 represents the results of hypothesis testing by comparing the pre- and post-test scores of both test classes.

Table 9  Hypothesis testing results

<table>
<thead>
<tr>
<th>Sig</th>
<th>Results</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>0.983</td>
<td>There is no significant difference in the HOTS abilities of students between the experimental and control groups.</td>
</tr>
<tr>
<td>Post-test</td>
<td>0.049</td>
<td>There is a significant difference in the HOTS abilities of students between the experimental and control groups.</td>
</tr>
</tbody>
</table>

Table 9 presents the results of the independent sample t-test conducted, comparing both test classes' pre-test and post-test scores. From these results, it can be concluded that in the initial testing phase, there was no significant difference between the two classes when looking at the pre-test scores. However, different results were found when examining the post-test scores of both classes. A significance value of 0.049, below the 5% threshold, indicates a significant difference between the experimental and control classes in terms of students' HOTS abilities.

The significant difference in post-test scores between the two classes suggests that using ChatGPT in physics education has a positive impact. Students' average higher-order thinking skills tend to improve in the experimental class that utilizes ChatGPT compared to the control class.

More specifically, a HOTS assessment instrument was employed to assess students' HOTS level. This instrument underwent testing by analyzing the test questions. Additionally, the formulated questions were designed to align with the levels of thinking C4-C6, referencing Bloom's taxonomy. Therefore, the test results obtained by students indicate they possess a strong proficiency in HOTS.

**Effect Size**

Effect size is a test conducted to measure the magnitude of the success of a study. The effectiveness of the PjBL (Project-based Learning) model compared to the use of ChatGPT is measured using the effect size formula. The testing results yield data as shown in Table 10.

Table 10  Hypothesis testing results

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>Data</th>
<th>Stdv</th>
<th>Efek Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eksp</td>
<td>76.80</td>
<td>20</td>
<td>8.63</td>
<td>0.64</td>
<td>Large</td>
</tr>
<tr>
<td>Contro</td>
<td>70.45</td>
<td>20</td>
<td>10.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10 presents the results of the effect size test conducted to measure the level of success between the PjBL learning model and the use of ChatGPT. Effect size is used to understand the extent of the difference between the two groups in terms of improving students' HOTS abilities. The table shows that in the experimental class using ChatGPT, the average improvement in students' HOTS abilities is 0.64, with a standard deviation of 8.63. Referring to Cohen’s d Effect Size Criteria Cohen’s d (Lee, 2000) to determine the magnitude of the effect size, this Effect Size falls into the "Large" category. This indicates that the use of ChatGPT significantly positively influences improving students' HOTS. The comparison of effect sizes between the two groups provides insight into the
fact that the use of ChatGPT in the experimental class contributes more significantly to enhancing students' HOTS abilities than in the control class.

The utilization of ChatGPT in the educational context is expanding, eliciting diverse responses (Levin et al., 2023). ChatGPT introduces a fresh perspective on how students can harness technology in their learning process. Previous literature has indicated how AI technologies like this can enhance education (Pham & Sampson, 2022). In education, improving various competencies, one of which is HOTS, has become a focus. HOTS skills have become increasingly essential for students today as they enable them to tackle complex challenges (Kurnia* et al., 2022). In this context, the findings of this research align with previous studies. For instance, Zajuli et al. (2019) revealed that PjBL is the most effective approach to enhancing students' HOTS abilities. Implementing information technology, including AI like ChatGPT, is also a crucial factor influencing HOTS skills.

The study found that students who participated in PjBL supported by ChatGPT achieved a higher average final score compared to those following conventional education, with scores of 76.8 compared to 70.45. Several factors within ChatGPT contributed to this improvement in students' abilities, including increased enthusiasm and activity among students when interacting with ChatGPT, the guidance and assistance it provided in comprehending complex physics concepts, the student's ability to construct robust scientific arguments, and a more precise focus on relevant literacy. This research provides strong evidence that the utilization of ChatGPT can significantly enhance the quality of physics education and encourage students to explore, create, and develop their understanding more boldly. The findings are reflected in the final results through HOTS tests, where students exhibited a notably significant increase compared to the control group, with an effect size of 0.64 (large). This result is further supported by Faiz & Kurniawaty's study (2023), which underscores the essential role of ChatGPT in education.

Another noteworthy finding is that students in the experimental class demonstrated more diverse performance outcomes during project implementation. They explored a wider array of tools provided by their teachers and exhibited greater creativity in executing physics projects. This contrasts with students in the control class, who tended to follow a more monotonous approach, often replicating projects without in-depth exploration. With the assistance of ChatGPT, students in the experimental class not only achieved improvements in HOTS scores but also displayed stronger critical thinking skills and a greater inclination to undertake unique physics projects. Dahlkemper et al. (2023) have indicated using ChatGPT across various learning contexts.

Overall, this research not only underscores the importance of technology in education but also emphasizes the need for appropriate teaching approaches and the imparting of relevant skills to confront future challenges. With the growing use of ChatGPT and AI technology in education, combining technology with effective teaching methods can help prepare students for an ever-evolving and changing world (Alqahtani et al., 2023). However, it is essential to remember that technology is just a tool, and the quality of teaching still depends on how teachers implement it and how students engage in the learning process. Therefore, successful education must integrate technology with effective learning strategies, enabling students to develop critical skills, conceptual understanding, and higher-order thinking skills. Hallal et al. (2023) acknowledge in their research
that AI still has many shortcomings, but it has great potential to enhance human intelligence. ChatGPT has made significant contributions in various sectors, including the fields of medicine, tourism, business, and, not least, education (Ali et al., 2023; Currie, 2023; Hunter et al., 2023; Koc et al., 2023). To achieve an innovative and effective future in education, it is crucial to continuously explore and harness AI technology’s potential. By leveraging this technology, the younger generation can be shaped into smarter individuals, equipped with relevant skills, and prepared to face the increasingly complex challenges in the digital age. AI support will be the key to realizing a brighter and more adaptive future in education.

CONCLUSION

In this research, the experimental and control groups were given HOTS questions as pre-tests to measure the initial abilities of the students. Post-tests were administered to observe changes in students' abilities after the learning intervention involving the implementation of ChatGPT and the Merdeka Belajar concept. The research results indicate a significant improvement in the HOTS abilities of the students in the experimental group after undergoing the intervention compared to the control group. Although the average pre-test scores between the two groups were not significantly different, there was a more noticeable increase in the average post-test scores for the experimental group (76.8) compared to the control group (70.45).

The frequency data categorization analysis also indicates intriguing results. ChatGPT in the experimental group helped some students reach the "Very High" category, which was previously unattained during the pre-test. These results suggest that using ChatGPT in the context of Merdeka Belajar can enhance students' HOTS abilities, especially in the higher value categories, and help reduce the number of students with very low abilities.

Statistical tests show that the data in this research is normally distributed and homogenous, ensuring the reliability of the analysis results. Hypothesis testing revealed a significant difference in improving students' HOTS abilities between the experimental and control groups, especially in the post-test scores. The calculated effect size indicates that the use of ChatGPT has a significantly positive and large effect on enhancing students' HOTS abilities, particularly in the experimental group.

Therefore, this research concludes that the use of ChatGPT in learning with the Merdeka Belajar approach has a significantly positive impact on improving students' HOTS abilities. The implications of this research encompass increased student engagement and creativity in developing projects during the learning process. The heightened student engagement reflects a positive response to the use of ChatGPT in education, where interaction with advanced technology positively impacts student participation and interest. Furthermore, the increased creativity in project development indicates that ChatGPT significantly motivates students to be more explorative and innovative in applying complex concepts. These implications serve as a positive indicator that the utilization of ChatGPT in the educational context can have a meaningful impact on the development of both cognitive and creative skills in students. These findings align with previous literature emphasizing the role of AI technology in enhancing learning and students' higher-order thinking abilities. It's important to consider that technology must be accompanied by the right teaching approach to better prepare students for future challenges.
For future research development, expanding the sample scope and including a wider variety of learning contexts is recommended. ChatGPT can be explored more deeply by integrating it into various subjects and learning contexts. Furthermore, research can involve more variables, such as student motivation, interactions between students and teachers, and the long-term impact of AI technology on learning. Additionally, the implementation of ChatGPT can be combined with other teaching methods to optimize student learning outcomes. By conducting more comprehensive research, we can gain deeper insights into the potential of AI technology utilization in education and how it can be effectively integrated into diverse learning environments.

REFERENCES


evaluasi higher order thinking skills menggunakan quizizz pada materi termokimia untuk meningkatkan kemampuan berpikir tingkat tinggi peserta didik. Jurnal Pendidikan Sains Indonesia, 10(1), 176–190. https://doi.org/10.24815/jpsi.v10i1.2


