Journal of Innovative Technology and Sustainability Education
Homepage : https://journal.i-ros.org/index.php/JITSE
Email : jitse@i-ros.org

e-ISSN 3109-2454
Vol. 1, No.3, November 2025

Bridging AI and Statistics: Comparative Evaluation of SPSS and ChatGPT in Educational Data Processing

Rahmatta Thoriq Lintangesukmanjaya[1*], Binar Kurnia Prahani[1], Dwikoranto[1]

[1]Universitas Negeri Surabaya, Surabaya, Indonesia

DOI : https://doi.org/10.63230/jitse.1.3.94

Sections Info

Article history:
Submitted: September 4, 2025
Final Revised: December 2, 2025
Accepted: December 2, 2025
Published: December 31 2025

Keywords:
AI;
Data;
Evaluation;
Statistics;
Technology.

ABSTRACT

Objective: *This study aims to compare the results of statistical data analysis using SPSS and Artificial Intelligence (AI), specifically ChatGPT, in processing students' learning outcomes. A quantitative experimental method was employed with a sample of 100 numerical data points representing students' academic performance.* ***Method:*** *This research is a comparative study, with processing outcomes in SPSS and ChatGPT.* ***Results:*** *Both SPSS and ChatGPT were utilized to perform statistical tests and analyze the same dataset. The findings indicate that both tools produced consistent results, demonstrating ChatGPT's potential as a complementary analytical tool. However, SPSS remains superior in terms of accuracy, academic legitimacy, and the availability of comprehensive features officially recognized in the research community. ChatGPT offers advantages in its simplicity, accessibility, and efficiency, particularly for quick analysis without software installation, making it a practical tool for educators and novice researchers.* ***Novelty:*** *In formal academic research contexts, SPSS remains the primary choice for ensuring methodological rigor and credibility, while AI-driven tools such as ChatGPT can serve as effective companions to support the analytical process and facilitate statistical learning. This study highlights the potential synergy between traditional statistical software and emerging AI technologies, emphasizing the importance of selecting tools that align with research needs, context, and academic standards. The findings provide insights into the evolving role of AI in research practices, suggesting that while AI enhances accessibility and user experience, conventional statistical software remains indispensable for high-stakes research validation.*

INTRODUCTION

With the rapid development of technology, especially as we enter the era of technology adoption, all activities are inextricably linked to its use, both digital and conventional. One frequently used technology is artificial intelligence (AI) (Sposato, 2025). The use of AI has now expanded into various sectors of life, from healthcare and education to business and everyday life (Venturini, 2025). This is due to AI's ability to execute and solve problems effectively. One example is in academia, where AI is often used in data retrieval, data analysis, and even report generation.

The importance of using AI lies in its ability to process large amounts of data quickly and accurately. AI can improve operational efficiency, support more informed decision-making, and automate processes that previously required human intervention. Furthermore, AI drives innovation and scientific development, opening new opportunities across various disciplines (Wan, 2025). One example is the implementation of AI in statistical testing. Artificial Intelligence (AI) applications, such as ChatGPT, are increasingly popular for data analysis compared to conventional statistical software such as SPSS, Anates, Winstep, and others. AI simplifies the analysis process because users do not require in-depth statistical technical skills (Chaudhry et al., 2023; Nikolic et al., 2023). Provide data and simple instructions.

The use of GPT in academia presents several challenges that require serious consideration. While this technology can help accelerate the writing process, organizing ideas, and understanding concepts, many academic institutions still question its reliability because GPT does not always produce precise information and can introduce bias or factual errors (Ferber et al., 2024). Another challenge is students' tendency to rely too heavily on AI to complete academic assignments, potentially diminishing the development of critical thinking, research literacy, and writing skills that should be developed independently. Furthermore, GPT lacks formal academic validity, making its interpretations or analyses unsuitable for use as primary references in scientific research. Although technology continues to advance (Lintangesukmanjaya et al., 2025), statistical evaluation of work must also be considered.

However, the accuracy of statistical tests using AI, especially ChatGPT, cannot be guaranteed. While Artificial Intelligence (AI) has many advantages in assisting with data analysis, decision-making, and report preparation, its use still has limitations and potential risks, particularly regarding accountability for the results. Furthermore, AI is highly dependent on the quality of user input and instructions (prompts). Therefore, the following is a simple study aimed at comparing statistical analysis results between AI (ChatGPT) and SPSS to determine the strengths and weaknesses of each.

RESEARCH METHOD

Types of research

This study, which compares the use of AI in SPSS and ChatGPT, falls under the category of comparative research, using a qualitative-quantitative (mixed-methods) approach. In this study, SPSS is used to analyze data, including quantitative validity and reliability tests. In contrast, ChatGPT is used as an artificial intelligence tool for qualitative analysis, result interpretation, and the development of scientific narratives. This study aims to assess the effectiveness, efficiency, and accuracy of each tool in supporting the data analysis process for educational or social research. Through this comparison, researchers can identify the advantages of SPSS as a numerical analysis tool for structural data and the strengths of ChatGPT in understanding context, suggesting interpretations, and producing reports that are communicative and adaptable to user needs.

The method used for the comparative analysis of statistical test results is a qualitative descriptive analysis of results processed in SPSS and ChatGPT. The data used is 20 in Excel/scv* format, which is artificial and obtained from statistics course sources. The following is the data used before analysis:

Population and Data Processing

Table 1. List of grades for class A and class B

NO	Kelas A		NO	Kelas B	
	Pre-test	***Post-test***		***Pre-test***	***Post-test***
1	65	70	1	65	70
2	72	75	2	79	80

NO	Kelas A		NO	Kelas B	
	Pre-test	*Post-test*		*Pre-test*	*Post-test*
3	65	75	3	78	82
4	78	80	4	72	75
5	66	75	5	69	75
6	70	90	6	72	74
7	80	82	7	82	84
8	72	72	8	68	78
9	76	84	9	70	70
10	60	85	10	65	68

The results were analyzed and compared between statistical tests using SPSS and ChatGPT. The analyses tested included prerequisite tests (normality and homogeneity tests) and difference tests (dependent and independent t-tests).

RESULTS AND DISCUSSION

Results

The results of the SPSS statistical test on 10 pre-test and post-test data sets showed an increase in average scores after the treatment. A comparison of the two sets of data also revealed a significant difference, as indicated by the statistical test results. Thus, these results confirm that the intervention had a positive impact on improving participants' abilities.

a. Pre-requisite Test

The first result obtained from the normality test is as follows,

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre_A	.153	10	.200*	.962	10	.806
Post_A	.223	10	.171	.950	10	.671

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre_B	.200	10	.200*	.921	10	.365
Post_B	.163	10	.200*	.919	10	.351

a) Class A **b)** Class B

Figure 1. Normality test SPSS

The results of the statistical test show that the significance value obtained in Shapiro-Wilk is > 0.05. Furthermore, the homogeneity results are obtained as follows:

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Pre	Based on Mean	.160	1	18	.694
	Based on Median	.150	1	18	.703
	Based on Median and with adjusted df	.150	1	17.991	.703
	Based on trimmed mean	.160	1	18	.694

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Post	Based on Mean	.682	1	18	.420
	Based on Median	.585	1	18	.454
	Based on Median and with adjusted df	.585	1	17.065	.455
	Based on trimmed mean	.681	1	18	.420

a) Class A **b)** Class B

Figure 2. Homogeneity test SPSS

The homogeneity results show homogeneous data with a sig. value > 0.05 so it is declared homogeneous.

b. Difference Test

1. T Dependent Test

A dependent T-test was conducted to determine the difference between the pre-test and post-test scores for each class. The results were as follows:

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Pre_A - Post_A	-8.400	8.208	2.596	-14.272	-2.528	-3.236	9	.010
Pair 2	Pre_B - Post_B	-3.600	2.875	.909	-5.657	-1.543	-3.959	9	.003

Figure 3. T-dependent results SPSS

2. T Independent Test

The independent T test was conducted to determine the differences in the values obtained from the two classes, namely class A and class B. In the results of this difference test, the results were carried out on each test as follows:

a) Pre-test

The results of 20 students from classes A and B were obtained as follows,

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Hasil	Equal variances assumed	.160	.694	-.580	18	.569	-1.60000	2.75762	-7.39354	4.19354
	Equal variances not assumed			-.580	17.861	.569	-1.60000	2.75762	-7.39677	4.19677

Figure 4. Results of the pre-test independent t-test SPSS

The results of the independent T-test obtained that the sig. (2-tailed) value > 0.05, namely 0.569, so it can be seen that there is no significant difference between learning outcomes in class A and class B.

b) Post-test

The results of 20 students from classes A and B are as follows,

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Hasil	Equal variances assumed	.692	.417	1.209	18	.242	3.20000	2.64575	-2.35852	8.75852
	Equal variances not assumed			1.209	17.475	.243	3.20000	2.64575	-2.37051	8.77051

Figure 5. SPSS Post-test t-independent test results

The results of the sig. (2-tailed) value > 0.05, which is 0.242, indicate that there is no significant difference between learning outcomes in class A and class B.

The ChatGPT statistical test results based on 10 pre-test and post-test scores showed a noticeable improvement in the mean values after the intervention. The comparison between the two data sets indicated a statistically significant difference, suggesting measurable learning progress. Overall, these findings support the conclusion that the applied treatment positively influenced participants' performance. GPT Chat cannot analyze complex excel data so in GPT chat analysis narrative data is used as shown below,

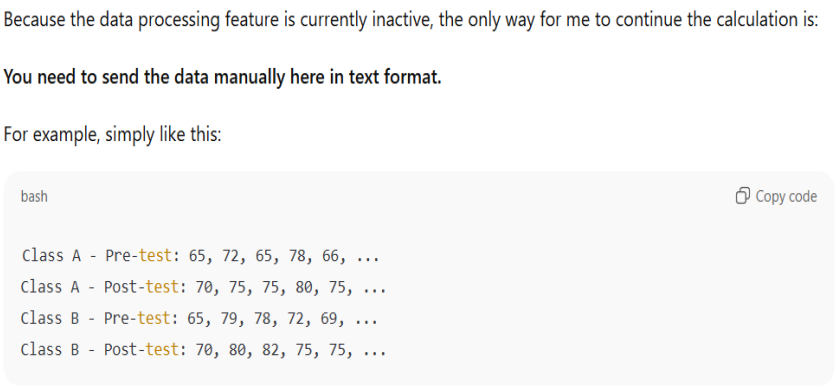

Figure 6. Prompt AI chat GPT

a. Pre-requisite Test

The first result obtained from the normality test is as follows,

Data	Shapiro-Wilk p-value	Kolmogorov-Smirnov p-value
Kelas A - Pre-test	0.8064	0.9104
Kelas A - Post-test	0.6707	0.5670
Kelas B - Pre-test	0.3647	0.7487
Kelas B - Post-test	0.7164	0.9131

Interpretasi singkat:

- Semua p-value > 0.05 → artinya **data berdistribusi normal** menurut kedua uji (baik Shapiro-Wilk maupun Kolmogorov-Smirnov).

Figure 7. GPT chat normality statistics results

The results of the statistical test show that the significance value obtained in the Shapiro-Wilk test is > 0.05. The following are the results of the homogeneity test as seen in the following figure:

Data	Levene Statistic	p-value	Kesimpulan
Pre-test	0.1500	0.7031	Varians homogen (p > 0.05)
Post-test	0.7101	0.4105	Varians homogen (p > 0.05)

Figure 8. GPT chat homogeneity statistics results

The results of the statistical test show that the significance value obtained p-value > 0.05 so it is declared homogeneous.

b. Difference Test

1. T Dependent Test

A dependent T-test was conducted to determine the difference between the pre-test and post-test scores for each class. The results were as follows:

Kelas	T-Statistic	p-value	Kesimpulan
Kelas A	-3.2361	0.0102	Terdapat perbedaan signifikan (p < 0.05)
Kelas B	-3.9595	0.0033	Terdapat perbedaan signifikan (p < 0.05)

Figure 9. Chat GPT dependent t-test statistical results

2. T Independent Test

The independent T test was conducted to determine the differences in the values obtained from the two classes, namely class A and class B. In the results of this difference test, the results were carried out on each test as follows:

Pengujian	T-Statistic	p-value	Kesimpulan
Pre-test	-0.5802	0.5690	Tidak ada perbedaan signifikan
Post-test	1.2095	0.2421	Tidak ada perbedaan signifikan

Figure 10. Chat GPT independent t-test statistical results

The results of the p-value > 0.05 show that there is no significant difference between learning outcomes in class A and class B, both pre-test and post-test scores.

Discussion

The results from each trial were generally similar, but several aspects and values listed in SPSS were not identified by the AI using ChatGPT. Normality tests revealed that both SPSS and ChatGPT AI results indicated that the data had a normality value >0.05,

thus indicating normality (Orcan, 2020). Because the data were identified as usual, parametric tests were used in both SPSS and ChatGPT. The results were similar, with significant differences in the dependent t-test and non-significant differences in the independent t-test. Although the results obtained from SPSS and ChatGPT were generally similar, specific differences in the output were evident. These differences were evident in the specifications, statistical analysis, and output quality of the two analyses.

SPSS has several advantages that make it one of the most popular statistical software programs in both academic and professional research. Its main strength lies in its comprehensive statistical features, ranging from fundamental analysis to advanced and multivariate tests (Ramdani et al., 2025; Salvador et al., 2024). Furthermore, SPSS's output tends to be more comprehensive, neat, and easy to read, making it easier for researchers to interpret the results. SPSS is also known for its high level of precision and allows for automated analysis requests (Huang et al., 2024). Furthermore, SPSS holds an official, academically recognized license, which enhances its credibility in scientific publications.

However, SPSS also has several drawbacks that need to be considered. The data input process in SPSS is considered more complex and less straightforward than some other statistical platforms (Attwall & Singh, 2024), so novice users may need time to adapt. Furthermore, SPSS is software that must be installed on specific devices and is not web-based, thus limiting the flexibility of use (Jain & Sengar). This installation requirement also often requires specific computer specifications and access to a paid license that is not always affordable for all users. Thus, while SPSS is powerful in terms of functionality, there are technical and accessibility barriers that can be challenging for some researchers or students.

ChatGPT has several advantages that make it easily accessible and highly practical for initial data analysis. One of its main advantages is its fast, simple data input process, which eliminates the need for users to navigate complex technical steps (Jumriah et al., 2024). Furthermore, ChatGPT does not require software installation, allowing it to be used directly through a browser at any time and on various devices. ChatGPT also excels in narrative flexibility—users can freely ask questions and customize explanations as needed, including crafting descriptions or interpreting data results in a specific language style. While they may appear similar at first glance, they differ significantly. This can be determined through a comparative analysis of statistical data using SPSS and ChatGPT (Mohammed, 2024). This comparison demonstrates that each option offers its own unique results and considerations. However, for academic purposes, SPSS remains recommended for more accurate data collection than ChatGPT analysis.

On the other hand, ChatGPT has several drawbacks when compared to specialized statistical software. Its statistical processing capabilities are limited because ChatGPT does not offer advanced automated statistical tests, and its results tend to be less comprehensive than software like SPSS. The model also lacks precision and cannot

automatically call or read variables like professional statistical applications (Hakiki et al., 2023). Furthermore, ChatGPT struggles to process complex statistical data and lacks academic licensing or official validation, making its analysis less suitable as a primary reference in scientific research. Therefore, ChatGPT is more suitable for use as an aid to interpretation and narrative development (Lingard, 2023), rather than as a primary statistical analysis tool.

Going forward, the use of SPSS in academia faces challenges due to the growing demand for statistical skills. Students and researchers are required not only to be proficient in operating this software (Rahayu et al., 2024) but also to have a deep understanding of statistical concepts to utilize its complex, advanced features. Furthermore, the need for installation, paid licensing, and devices with specific specifications poses significant obstacles for both institutions and individuals. The biggest challenge is how universities can provide adequate training, access, and technical support to optimize SPSS use and prevent it from becoming a tool used mechanically without proper analytical understanding.

Meanwhile, the use of ChatGPT in academic contexts also presents new, equally significant challenges. While this AI helps expedite the writing process, initial data interpretation, and information retrieval, its limitations in producing precise statistical analyses are a significant concern. ChatGPT also has the potential for misuse, such as producing less original work or failing to adhere to academic validity principles. Furthermore, the lack of scientific licensing and formal verification standards makes AI analysis results unsuitable for primary academic reference. The challenge ahead is how educational institutions can guide the ethical, critical, and proportionate use of ChatGPT, ensuring this technology serves as a supporting tool that enriches digital literacy without compromising academic quality and integrity.

CONCLUSION

Fundamental Findings: The comparative analysis showed that SPSS and ChatGPT generally produced similar statistical results, particularly in normality testing and fundamental parametric analyses, although SPSS provided more detailed and precise outputs. SPSS demonstrated clear advantages in analytical depth, accuracy, and academic credibility, whereas ChatGPT offered accessibility, ease of use, and flexibility in narrative interpretation. Overall, the findings indicate that while ChatGPT is beneficial for preliminary analysis and explanatory support, SPSS remains the more reliable tool for rigorous academic statistical research. **Implication:** Based on the comparative analysis, the statistical test results from SPSS and the AI-based ChatGPT yielded similar results. However, SPSS excels in accuracy, academic validity, and the comprehensiveness of officially recognized analysis features in the research world. **Limitation:** Meanwhile, using AI, ChatGPT is very helpful for simple and fast data input, requiring no software installation. **Novelty:** For formal academic research, SPSS remains the primary choice, while ChatGPT can be an effective companion to simplify statistical analysis and learning.

ACKNOWLEDGEMENTS
The researcher would like to thank the teachers and students of SMA Negeri 19 Surabaya who helped during the research until its completion.

AUTHOR CONTRIBUTIONS
Rahmatta Thoriq Lintangesukmanjaya contributed to the conceptual framework, research design, and validation process; **Binar Kurnia Prahani** was involved in methodology development, data analysis, sourcing references, and drafting the manuscript; **Dwikoranto** handled data management, project coordination.

CONFLICT OF INTEREST STATEMENT
There are no conflicts of interest.

ETHICAL COMPLIANCE STATEMENT
This manuscript complies with research and publication ethics. The authors affirm that the work is original, conducted with academic integrity, and free from any unethical practices, including plagiarism.

STATEMENT ON THE USE OF AI OR DIGITAL TOOLS IN WRITING
The authors acknowledge the use of digital tools, including AI-based technologies, as support in the research and writing stages of this article. Specifically, Grammarly was employed for generated and paraphrase. All outputs generated with digital assistance were critically evaluated and revised to ensure academic rigor and ethical standards were upheld. The final responsibility for the manuscript rests entirely with the authors.

REFERENCES
Attwal, K., & Singh, A. (2024). Exploring SPSS statistics for data mining and statistical modeling. *International Journal of Statistics and Applied Mathematics*, *9*(3), 09-16. https://doi.org/10.22271/maths.2024.v9.i3a.1721

Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to revisit existing student's performance evaluation approach in higher education sector in a new era of ChatGPT a case study. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186X.2023.2210461

Ferber, D., Wiest, I. C., Wölflein, G., Ebert, M. P., Beutel, G., Eckardt, J. N., ... & Kather, J. N. (2024). GPT-4 for information retrieval and comparison of medical oncology guidelines. *Nejm Ai*, *1*(6), AIcs2300235. DOI: 10.1056/AIcs2300235

Hakiki, M., Fadli, R., Samala, A. D., Fricticarani, A., Dayurni, P., Rahmadani, K., ... & Sabir, A. (2023). Exploring the impact of using Chat-GPT on student learning outcomes in technology learning: The comprehensive experiment. *Advances in Mobile Learning Educational Research*, *3*(2), 859-872. https://doi.org/10.25082/AMLER.2023.02.013

Huang, Y., Wu, R., He, J., & Xiang, Y. (2024). Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: A comparative analysis with SAS, SPSS, and R. *Journal of Global Health*, *14*, 04070. https://doi.org/10.7189/jogh.14.04070

Jain, P., & Sengar, S. (2024). Unraveling The Role Of IBM SPSS: A comprehensive examination of usage patterns, perceived benefits, and challenges in research practice. *Educational Administration: Theory and Practice*, *30*(5), 9523-9530. https://kuey.net/

Jumriah, J., Susilawati, E. S., Supriatna, E., Smas, M. H., & Arini, I. (2024). Analysis of the use of GPT chat to improve student performance. *EDUKASIA Jurnal Pendidikan dan Pembelajaran*, *5*(1), 127-136. https://doi.org/10.62775/edukasia.v5i1.725

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on medical education*, *12*(1), 261. https://doi.org/10.5334/pme.1072

Lintangesukmanjaya, R. T., Prahani, B. K., Dwikoranto, D., Alhusni, H. Z., & Kurtuluş, M. A. (2025). Technology integration for SDGs-oriented social science education: A bibliometric perspective. In *E3S Web of Conferences*, *640,* 02017. https://doi.org/10.1051/e3sconf/202564002017

Mohamed, A. M. F. S. (2024). A comparative evaluation of statistical product and service solutions (SPSS) and ChatGPT-4 in statistical analyses. *Cureus*, *16*(10).

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, *48*(4), 559–614. https://doi.org/10.1080/03043797.2023.2213169

Orcan, F. (2020). Parametric or non-parametric: Skewness to test normality for mean comparison. *International Journal of Assessment Tools in Education*, *7*(2), 255–265. https://doi.org/10.21449/ijate.656077

Rahayu, N. I., Muktiarni, M., & Hidayat, Y. (2024). An application of statistical testing: A guide to basic parametric statistics in educational research using SPSS. *ASEAN Journal of Science and Engineering*, *4*(3), 569-582.

Ramdani, F., Hilmiyah, F., & Indriyani, V. (2025). The impact of SPSS on research completion. *TOFEDU: The Future of Education Journal*, *4*(2), 419-429. https://doi.org/10.61445/tofedu.v4i2.338

Salvador, M., de Jesus, L. S., & Magno, M. (2024). Economy, development, and fishermen's income in Atauro Island affected by natural disaster impacts. *Asian Journal of Engineering, Social and Health*, *3*(11), 2529-2547. https://doi.org/10.46799/ajesh.v3i11.457

Sposato, M. (2025). Artificial intelligence in educational leadership: a comprehensive taxonomy and future directions. *International Journal of Educational Technology in Higher Education*, *22*(1), 1–18. https://doi.org/10.1186/s41239-025-00517-1

Venturini, R. E. (2025). *Technological innovations in agriculture : the application of Blockchain and Artificial Intelligence for grain traceability and protection Inovações tecnológicas na*

agricultura : aplicação de Blockchain e Inteligência Artificial para a rastreabilidade e p. 1–9. https://doi.org/10.34117/bjdv11n3-007

Wan, Y. (2025). *Artificial Intelligence Empowered Reforms in Economics Education*. *January*. https://doi.org/10.5281/zenodo.14864252

***Rahmatta Thoriq Lintangesukmanjaya (Corresponding Author)**
Postgraduate Physics Education, Universitas Negeri Surabaya
Jl. Ketintang, Ketintang, Kec. Gayungan, Surabaya, Jawa Timur 60231
24031635011@mhs.unesa.ac.id

Binar Kurnia Prahani
Physics Education, Universitas Negeri Surabaya
Jl. Ketintang, Ketintang, Kec. Gayungan, Surabaya, Jawa Timur 60231
binraprahani@unesa.ac.id

Dwikoranto
Postgraduate Physics Education, Universitas Negeri Surabaya
Jl. Ketintang, Ketintang, Kec. Gayungan, Surabaya, Jawa Timur 60231
dwikoranto@unesa.ac.id