Assessing text comprehension proficiency: Indonesian higher education students vs ChatGPT

Juanda – Iswan Afandi

DOI: 10.18355/XL.2024.17.01.04

Abstract

AI has developed rapidly. However, AI research in applied linguistics in the field of language education in Indonesia still needs to be expanded to reading and writing skills. This research aims to explore students' skills in writing summaries and understanding the historical theme of the development of the Indonesian language with AI based on gender and university aspects. The quantitative method uses descriptive statistical analysis techniques, independent sample t-test, and Welch One-Way ANOVA. The research sample was 288 students from Makassar State University, Timor University, and Makassar Health Polytechnic. The results show that ChatGPT is significantly better at reprocessing text than students based on the aspects measured. ChatGPT outperforms almost every aspect of the assessment. However, in the MCT_Score aspect, the average for Universitas Negeri Makassar students is slightly higher than ChatGPT and the other two universities. Meanwhile, the Makassar Health Polytechnic almost matches the average ChatGPT score. Apart from that, the Universitas Timor average seems significantly different, with a score range of only 6.00 – 7.00. This research contributes to developing the Indonesian curriculum using Artificial Intelligence (AI) technology. The government can use these findings as a basis for making better policies to improve the quality of education. This research implies that Indonesian students have a gap in understanding texts compared to ChatGPT. The first implication is the need to revise and develop the educational curriculum. Therefore, future research can examine text comprehension abilities in more specific contexts, such as scientific texts, journalism, literature, or specific scientific disciplines. It can provide more detailed insight into students' strategies for overcoming difficulties in understanding texts. In addition, future research will be conducted on the broader impact of using Artificial Intelligence technology in language education on the development of student text comprehension and the potential social and ethical impacts.

Key words: Artificial Intelligence, ChatGPT, language education, higher education students, text comprehension, writing skills

Introduction

Language as a communication tool is unique in acquiring and learning from one individual to another. As language users, humans have competed with AI in language acquisition, mastery, and aspects of learning methods. In contemporary times, educators harness advanced technical and informational tools to craft engaging learning environments. At the same time, students can access educational materials from anywhere at any time through their devices. Innovative technologies facilitate individual and collaborative learning, skill development, competency training, and diversifying educational content (Lukianenko & Vadaska, 2020). In Ukraine, the way to continue education in war conditions found widespread adoption of online and blended teaching models. The ongoing conflict significantly disrupts the regular operations of schools and educational institutions, impeding students' access to a secure and stable learning environment. The need for more resources, such as books, equipment, and skilled teachers, worsens the educational system's challenges.

Furthermore, the continual menace of violence and displacement further obstruct students' capacity to attend school consistently. Despite these challenges, local and international organizations are working to offer educational support and resources to alleviate the impact of the conflict on the education of Ukrainian children and youth. This global trend is on the rise and has garnered support from the United Nations, which has advocated for a shift in teaching methods. The digital revolution is considered the cornerstone of this transformation. Reports from 2022 indicate that effective utilization of connectivity and openly accessible digital educational resources has the potential to play a significant role in the transformation and democratization of education (Baklazhenko & Kornieva, 2023).

Research related to the comparison of text reading comprehension results between humans and AI has been carried out by several researchers (Dao, 2023; Desaire et al., 2023; Duenas et al., 2023; Giannos & Delardas, 2023; Vázquez-Cano et al., 2023; Xiao et al., 2023). The research presents a study that evaluates the scores obtained by ChatGPT when summarizing reading comprehension texts from the PISA international test with prompts that made him perform a simulation as if he were a 15-year-old student. The research found that ChatGPT summaries scored best in content and writing style, with scores 3 and 2.5 points higher than student scores (Vázquez-Cano et al., 2023). This study compares academic texts authored by academic scientists and those generated by ChatGPT. The findings indicate that humans favor a broader range of sentence structures than ChatGPT. While the average sentence length does not significantly distinguish between the two groups, distinctive factors include the standard deviation of sentence lengths within a specific paragraph and the median difference in word count between individual and subsequent sentences (Desaire et al., 2023). A separate research investigation analyzed a reading comprehension system designed to offer tailored and top-tier reading materials to middle school English students in China. A comprehensive assessment of the generated reading materials and their accompanying practice questions, both through automated and manual means, reveals that the content produced by the system is well-suited for students and outperforms the quality of presently available human-authored materials (Xiao et al., 2023).

The research examines the comparative performance of three large language models (LLM), namely OpenAI ChatGPT, Microsoft Bing Chat (BingChat), and Google Bard, on the VietNamese High School Graduation Examination (VNHSGE) English language dataset. The performance of BingChat, Bard, and ChatGPT (GPT-3.5) was 92.4%, 86%, and 79.2%, respectively. The results show that BingChat is better than ChatGPT and Bard. The results show that BingChat, Bard, and ChatGPT outperform Vietnamese students in English language proficiency (Dao, 2023). The research examines the comparison of creating multiple choice items for reading comprehension by ChatGPT with those created by humans. The research results show that ChatGPT can produce explanations with various types of information comparable to those created by humans (Duenas et al., 2023).

Additionally, a research study was undertaken to evaluate ChatGPT's performance in conventional admission tests within the UK, including assessments like the BioMedical Admissions Test (BMAT), Test of Mathematics for University Admission (TMUA), Law National Aptitude Test (LNAT), and Thinking Skills Assessment (TSA). The aim was to gauge its potential as an innovative tool for educational support and test preparation. The results reveal that the proportion of correct responses is significantly lower compared to incorrect ones (Giannos & Delardas, 2023).

AI research in Indonesia in linguistics has been carried out, including the work by Yudono (2023) examining AI's ability to write short horror stories. This study has identified several limitations in AI-generated horror short stories, including challenges in composing titles, crafting realistic dialogues, incorporating intrinsic story elements, and utilizing olfactory imagery. AI does not generate story titles, formulate direct and engaging dialogues, reiterate essential story backgrounds, or employ the language style associated with olfactory imagery in its storytelling. Research (Prastiwi & Pujiawati, 2019) examines the combination of Artificial Intelligence (Paperrater) with natural intelligence in learning to write English. Through Paperrater, students can create better written English compositions; through natural intelligence, students can distinguish which feedback needs to be corrected and which should be ignored. Abimanto and Mahendro's research (2023) examines the effectiveness of using artificial intelligence (AI) in English language learning. The study revealed substantial enhancements in listening, speaking, reading, and writing proficiencies following the utilization of AI. These results support the efficacy of integrating AI technology into language learning practices.

According to previous research, ChatGPT had higher content and writing style ratings than students. At the same time, significant differences in sentence structure were found between human-generated and ChatGPT-generated texts. Additionally, ChatGPT proficiently created explanations parallel to those generated by humans to create multiple-choice questions about reading comprehension. ChatGPT demonstrated advantages in certain areas, including text summarization and the creation of comprehension exercises. Therefore, weighing the pros and cons of implementing AI technologies in education is crucial. Based on several previous studies, it turns out that no research has been found that focuses on comparing the text comprehension results of bachelor students in Indonesia based on gender and university and comparing the text comprehension results between bachelor students and ChatGPT model GPT-3.5. The problem with this research is that there needs to be a clear understanding of the level of understanding between AI ChatGPT model GPT-3.5 and students. Therefore, this study aims to compare text comprehension between university students and ChatGPT through various statistical tests. The researcher formulated the research hypothesis as follows:

H1a: Significant differences exist in text comprehension results based on gender groups.
H1b: There are significant differences in text comprehension results based on university groups.
H2: ChatGPT has higher text comprehension results than students.

This research contributes to developing the Indonesian curriculum by utilizing technology such as Artificial Intelligence (AI). This research can help assess the extent to which the education curriculum in Indonesia has succeeded in developing tertiary students' text comprehension abilities. By comparing students' text comprehension abilities with ChatGPT, this research can provide insight into whether the existing curriculum is adequate in developing these abilities or needs changes.

Literature Review

AI literacy

AI literacy refers to an individual's understanding and ability to recognize, comprehend, and interact with artificial intelligence (AI). It includes knowledge of how AI works, its types, and its impact on various aspects of life (Laupichler et al., 2022, 2023). AI literacy involves skills that can be applied to increase efficiency and innovation in various industries (Carolus et al., 2023; Hornberger et al., 2023). Individuals who have physical access to information and communication technology are more expected to use and recognize AI (Celik, 2023, p. 1). With the increasing role of AI in modern society, AI literacy is becoming increasingly important, both for the general public and professionals, so that they make better decisions about the use of AI, avoid misuse of AI, and contribute to the development of this technology. Individual factors reflect AI competencies: technical understanding, critical judgment, and practical application (Laupichler et al., 2023: 1).

The development of AI literacy helps reduce the digital gap between individuals and groups with access to and understanding AI and those without (S.-C. Kong et al., 2022, p. 1). It allows more people to participate in the development and use of AI and promotes inclusion in the AI era. Using social robots as learning companions has been proven to help students understand AI principles (Su & Zhong, 2022: 1). Participating students could propose authentic scenarios, apply their new knowledge of AI, and devise meaningful AI-based solutions in digital stories (Mertala et al., 2022; Ng et al., 2022). Therefore, AI literacy education and training must prepare society to face a future increasingly connected to AI technology (Dai et al., 2023: 84).

AI in Language Education

Artificial intelligence (AI) has significantly changed the landscape of language education. AI in language education enables a more personalized and practical learning experience. Through massive analysis of student data, AI can identify weaknesses in language comprehension, measure speaking and writing levels, and design a customized curriculum for each individual. To seamlessly integrate and leverage key language models within the educational framework and teaching curriculum, it is imperative to establish a well-defined strategy within the education system. Adopting a straightforward pedagogical approach that substantially emphasizes critical thinking and fact-checking strategies is essential (Kasneci et al., 2023). AI is a tool that has the potential to help language learners process language in a more structured way than traditional word processors (Gayed et al., 2022; González-Lloret, 2023).

Applications of AI within academics and education encompass various domains, such as providing educational assistance and constructive feedback, facilitating assessments, tailoring curricula to individual needs, offering personalized career guidance, and delivering support for mental health and well-being (Alqahtani et al., 2023; Lim et al., 2023). With the help of chatbots and virtual assistants, students can practice speaking and writing languages interactively, which helps improve their confidence and communication skills. Moreover, AI empowers educators to precisely gauge student advancement and offer prompt feedback, thereby enhancing the efficiency and effectiveness of language education. Nevertheless, there exist both challenges and prospects concerning AI literacy in education, specifically: (1) a deficiency in educators' knowledge, skills, and confidence regarding AI; (2) inadequacies in curriculum design; and (3) a shortage of established teaching guidelines (Su et al., 2023, p. 1). Consequently, the University of Florida (UF) integrates AI into its curriculum. It creates avenues for student involvement in identified facets of AI literacy, irrespective of the student's field of study (Southworth et al., 2023, p. 1).

AI facilitates more accessible and cost-effective availability of language learning resources. AI-driven learning methodologies can enhance the advantages of personalized AI support and reinforce the sense of a collaborative learning environment between humans and AI (Gill et al., 2024; Wang, Liu et al., 2023). AI-powered language learning app, students can learn anytime and anywhere, without limitations of time or place. Utilizing GPT for Automatic Essay Scoring (AES) offers accuracy and reliability, contributing valuable assistance to human evaluators (Mizumoto & Eguchi, 2023, p. 1).

The latest technology enables better automatic translation, making language learning materials from various languages more accessible. Additionally, AI helps teach languages to people with disabilities with voice or text-based tools, opening up more inclusive educational opportunities. Based on semi-structured interviews conducted with twelve instructors at a higher education institution in Hong Kong, the findings highlight the importance of gaining competence and self-assurance in utilizing AI-based teaching tools. They also shed light on the difficulties and apprehensions encountered by language instructors and underscore the demand for tailored support in this context (Kohnke et al., 2023, p. 1). In this way, AI brings significant innovation to language education, helping students learn more effectively, efficiently, and inclusively. The readiness to employ AI in education, encompassing cognition, capability, and a forward-looking vision, is positively associated with ethical considerations. Each of the four components of AI readiness demonstrates favorable correlations, whereas the perceived threat posed by AI exhibits a negative correlation. Conversely, AI-facilitated innovation positively correlates with teacher job satisfaction (Ramadevi et al., 2023; Wang, Li, et al., 2023).

ChatGPT

ChatGPT is a language model developed by OpenAI based on the GPT-3.5 architecture. ChatGPT is designed to perform text-based tasks such as talking to users, answering questions, and interacting naturally in human language. This model has been trained with various text sources from the internet so that it has extensive knowledge until 2021. ChatGPT can be used in various contexts, from virtual assistants for customer support and online tutors to writing and communication tools. ChatGPT can establish interactive learning environments and replicate genuine engineering thought processes (Kong et al., 2023: 1). The increasing prominence of ChatGPT, an advanced language model that employs deep learning techniques to simulate human-like conversations, has raised concerns about its potential for misuse, particularly within the context of academic environments (Sweeney, 2023; Tsai et al., 2023). Honesty and humility have the most substantial relationship with the intention to use chatbot-generated text to commit academic cheating (Greitemeyer & Kastenmüller, 2023: 1).

One of the exciting features of ChatGPT is its ability to understand the context of a conversation and generate relevant responses (Choi et al., 2023). This model can adapt to various languages and conversational styles, making it useful in various communication applications. However, like all AI technology, ChatGPT has limitations, including a tendency to produce information that is not always accurate or produce inappropriate content in some cases. Therefore, the use of ChatGPT needs to be managed wisely and monitored to ensure ethical and beneficial use in various contexts (Dalalah & Dalalah, 2023: 1).

Methods

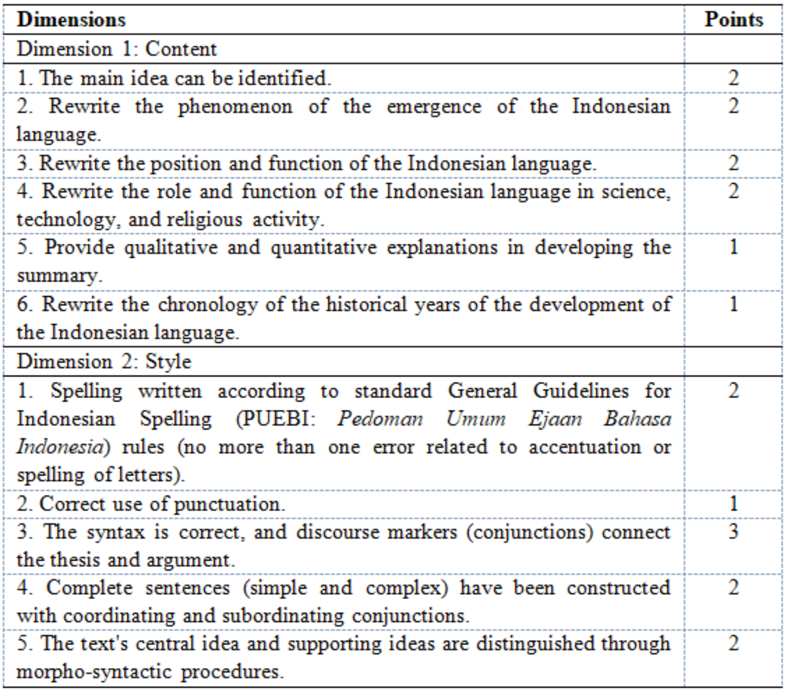
Research design

This research uses quantitative descriptive analysis to compare the results of students' understanding of historical texts on the development of the Indonesian language and the AI ChatGPT technology GPT-3.5 model. The research was carried out in two broad stages: research on students and AI (ChatGPT). In the first stage, the researcher selected a sample of students who read the text. After that, students were given a comprehension test consisting of multiple-choice and summary tests.

In the second stage, researchers input text as AI learning material via ChatGPT model GPT-3.5. Afterward, the researcher input the instructions: "Based on several pieces of text that have been given, summarize around 250-300 words." ChatGPT processes the text into memory into as many summaries as desired. Then, a multiple-choice test with ten questions is given. ChatGPT answers these questions based on text input provided by the researcher.

Finally, to ensure that the evaluation is homogeneous, the researcher created a summary assessment rubric according to two aspects, namely content and writing style, by modifying several assessment aspects that had been developed previously (Vázquez-Cano et al., 2023).

Table 1. Assessment rubric



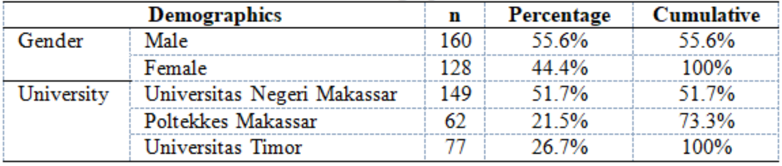
Dimensions	Points
Dimension 1: Content	
1. The main idea can be identified.	2
2. Rewrite the phenomenon of the emergence of the Indonesian language.	2
3. Rewrite the position and function of the Indonesian language.	2
4. Rewrite the role and function of the Indonesian language in science, technology, and religious activity.	2
5. Provide qualitative and quantitative explanations in developing the summary.	1
6. Rewrite the chronology of the historical years of the development of the Indonesian language.	1
Dimension 2: Style	
1. Spelling written according to standard General Guidelines for Indonesian Spelling (PUEBI: *Pedoman Umum Ejaan Bahasa Indonesia*) rules (no more than one error related to accentuation or spelling of letters).	2
2. Correct use of punctuation.	1
3. The syntax is correct, and discourse markers (conjunctions) connect the thesis and argument.	3
4. Complete sentences (simple and complex) have been constructed with coordinating and subordinating conjunctions.	2
5. The text's central idea and supporting ideas are distinguished through morpho-syntactic procedures.	2

Population and sample

The population of this study was N=1034 students from Universitas Negeri Makassar (UNM), Universitas Timor (Unimor, and Makassar Health Polytechnic (Poltekkes Makassar). The number of samples was n=288 students obtained through simple random sampling techniques. The number of samples is calculated based on Slovin's formula as follows.

$$n = \frac{N}{1+Ne^2} \qquad (1)$$

Table 2. Demographic Data



Demographics		n	Percentage	Cumulative
Gender	Male	160	55.6%	55.6%
	Female	128	44.4%	100%
University	Universitas Negeri Makassar	149	51.7%	51.7%
	Poltekkes Makassar	62	21.5%	73.3%
	Universitas Timor	77	26.7%	100%

Instruments, procedures, and data collection

Researchers use tests as research instruments. The instrument consists of a summarizing test and multiple choice. Before the test, the researcher gave students a text entitled "History of the Development of the Indonesian Language," which contained 4574 words. The text was obtained from the Indonesian language module of the Open University of Indonesia (Pramuki, 2018). The text comprehension test was carried out online using the help of a Google Form, which was distributed to students from selected universities. Data collection will be carried out in September 2023.

Data analysis

This research uses descriptive and inferential statistical analysis techniques. Descriptive statistical analysis calculated the difference between students' mean scores and ChatGPT. The inferential analysis method used in this research is the independent sample t-test and Welch ANOVA. Independent sample t-test was used to determine differences in scores based on gender groups, namely men and women. Meanwhile, Welch One-Way ANOVA was used to determine differences in student scores based on university. The normality and homogeneity tests are assumption tests carried out in this research. The Q-Q Plot normality test was carried out to test the normality of the data distribution. Meanwhile, Levene's test was carried out to test the homogeneity of variants. Researchers used the Jamovi 2.3.28 program to carry out these statistical tests.

Results

Descriptive statistics

Table 3 contains descriptive statistics for the five variables mentioned: MCT (Multiple Choice Test), SMT_Content (Summarizing Test-Content), SMT_Style (Summarizing Test-Style), SMT (Summarizing Test Total Score), and Overall_Score. This table contains important information about the samples used in the analysis. This sample consists of 289 observations (including ChatGPT), and no missing data in any variable indicates good data integrity.

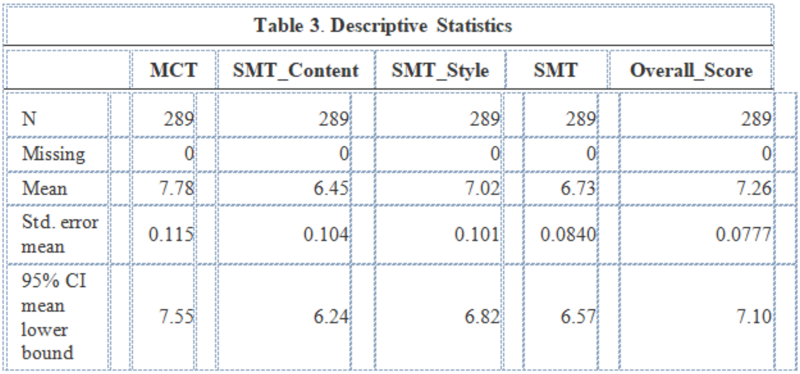
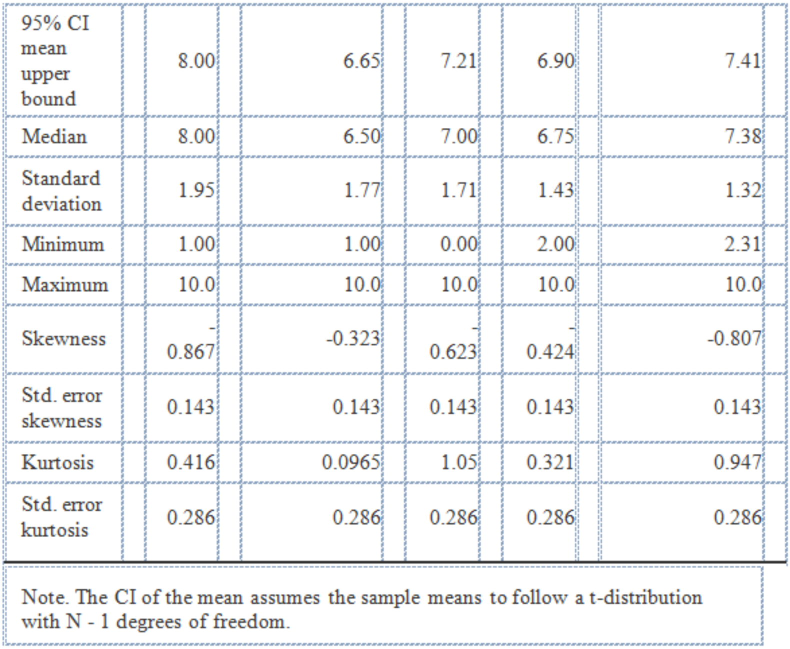


Table 3. Descriptive Statistics

	MCT	SMT_Content	SMT_Style	SMT	Overall_Score
N	289	289	289	289	289
Missing	0	0	0	0	0
Mean	7.78	6.45	7.02	6.73	7.26
Std. error mean	0.115	0.104	0.101	0.0840	0.0777
95% CI mean lower bound	7.55	6.24	6.82	6.57	7.10



95% CI mean upper bound	8.00	6.65	7.21	6.90	7.41
Median	8.00	6.50	7.00	6.75	7.38
Standard deviation	1.95	1.77	1.71	1.43	1.32
Minimum	1.00	1.00	0.00	2.00	2.31
Maximum	10.0	10.0	10.0	10.0	10.0
Skewness	-0.867	-0.323	-0.623	-0.424	-0.807
Std. error skewness	0.143	0.143	0.143	0.143	0.143
Kurtosis	0.416	0.0965	1.05	0.321	0.947
Std. error kurtosis	0.286	0.286	0.286	0.286	0.286

Note. The CI of the mean assumes the sample means to follow a t-distribution with N - 1 degrees of freedom.

Mean MCT was 7.78, whereas SMT_Content has a mean of 6.45, SMT_Style around 7.02, SMT 6.73, and Overall_Score has a mean of 7.26. Standard error of the mean (standard mean) relative error small For all variables, denotes that the mean Enough can reliable. Next, a 95% confidence interval for the mean is provided. We limit the bottom and limit The estimated range around the mean. The median of these variables has also been included, depicting the data distribution's middle value. Standard deviation indicates the degree of variation in the data. It can be seen that MCT has the highest standard deviation of around 1.95, while SMT_Content has a standard deviation of around 1.77. The data range, namely the minimum and maximum values, shows that all variables have a scale of 1 to 10. Skewness describes the extent to which the data distribution is asymmetrical, with negative values indicating the presence of negative skewness in the MCT, SMT_Style, SMT, and Overall_Score distributions. Kurtosis measures the degree to which a data distribution is more or less likely to be conical compared to a normal distribution. The kurtosis results show that all variables have positive kurtosis.

Comparison between students

An Independent sample t-test was conducted to evaluate significant differences in text comprehension results between men and women. To better understand these possible differences, this study looked at the p-value and effect size. The analysis results are presented in Table 4 below, which provides a clear insight into potential differences in text comprehension between the two gender groups.

Table 4. Score Comparison Based on Gender

	Student's t	df	p	Mean difference	SE difference	Effect Size
MCT	-0.148	286	0.882	-0.0344	0.232	-0.0176
SMT_Content	1.372	286	0.171	0.2883	0.210	0.1627
SMT_Style	1,089	286	0.277	0.2202	0.202	0.1291
SMT	1.504	286	0.134	0.2537	0.169	0.1783
Overall_Score	0.702	286	0.483	0.1100	0.157	0.0832

Note. H_a $\mu_{Female} \neq \mu_{Male}$

A vital consideration in this analysis is the p-value, an indicator of statistical significance. In this context, the null hypothesis (H0) states that no significant difference exists between men and women in text comprehension. In this research, the significance level generally used is 0.05. The analysis results show that the p-value for all these variables exceeds the specified significance level. Therefore, based on this analysis, there is insufficient statistical evidence to reject the null hypothesis. Based on existing data, there is no significant difference between male and female groups in text comprehension.

In addition, the effect size was measured using Cohen's d. The effect size provides an idea of the extent of the difference between two groups in standard units. Cohen's d values of all variables are relatively small in this context. It shows that if there are differences, they do not substantially impact text comprehension between men and women. The validity of a hypothesis can be accepted if the data meets the assumption test. Figure 1 below shows the level of normality of data using a Q-Q plot.

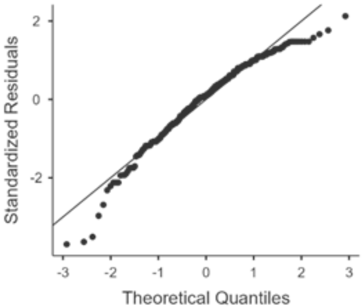
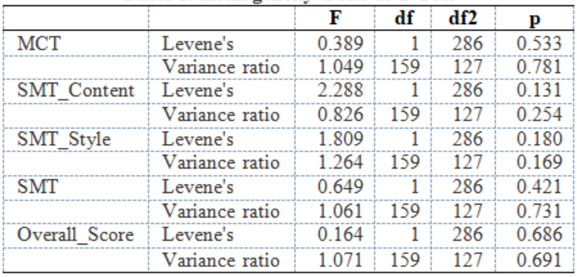


Figure 1: Normality Q-Q Plot of T-Test

Figure 1 above shows a typical data distribution. It is shown by the tendency of the data plot to approach a linear line. Even though the tail still looks far away, the distance between the plot and the linear line is insignificant. Therefore, the normality of data distribution is acceptable. Apart from normality, Table 5 below presents a homogeneity test using the Levene value.

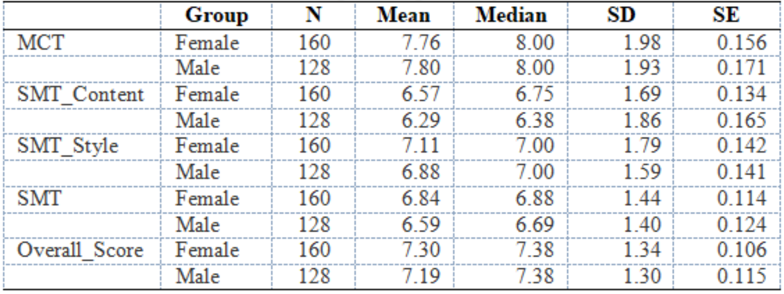
Table 5. Homogeneity Tests of T-Test



		F	df	df2	p
MCT	Levene's	0.389	1	286	0.533
	Variance ratio	1.049	159	127	0.781
SMT_Content	Levene's	2.288	1	286	0.131
	Variance ratio	0.826	159	127	0.254
SMT_Style	Levene's	1.809	1	286	0.180
	Variance ratio	1.264	159	127	0.169
SMT	Levene's	0.649	1	286	0.421
	Variance ratio	1.061	159	127	0.731
Overall_Score	Levene's	0.164	1	286	0.686
	Variance ratio	1.071	159	127	0.691

Levene's test is used to test whether the variance of the groups compared in the study is homogeneous. Table 5 above shows that the variance of all variables is not significantly different. The p-value (significance) is more significant than the alpha level used (0.05), which indicates that the assumption of homogeneity of variance is met. In other words, the variances of these groups are relatively equal.

Table 6. Group Descriptives by Gender



	Group	N	Mean	Median	SD	SE
MCT	Female	160	7.76	8.00	1.98	0.156
	Male	128	7.80	8.00	1.93	0.171
SMT_Content	Female	160	6.57	6.75	1.69	0.134
	Male	128	6.29	6.38	1.86	0.165
SMT_Style	Female	160	7.11	7.00	1.79	0.142
	Male	128	6.88	7.00	1.59	0.141
SMT	Female	160	6.84	6.88	1.44	0.114
	Male	128	6.59	6.69	1.40	0.124
Overall_Score	Female	160	7.30	7.38	1.34	0.106
	Male	128	7.19	7.38	1.30	0.115

Table 6 above provides information about the differences in scores between men and women in several tests. The average Multiple-Choice Test (MCT) score for men and women has very little difference, with the average for men being 7.80 and women 7.76. It shows no significant difference between the text comprehension results of men and women because the average difference is minimal.

However, when looking at other variables, such as Summarizing Test-Content Score (SMT_Content) and Summarizing Test-Style Score (SMT_Style), it can be seen that women tend to have a higher average score than men. Regarding content understanding and summary writing style, women may have slightly better abilities than men. Furthermore, considering the Summarizing Test Total (SMT) and Overall Score scores, the differences between men and women are also very small. In this case, there is no significant difference between the two.
Therefore, the differences between men and women appear to be most pronounced in SMT_Content and SMT_Style, with women having slightly higher averages. However, this difference is not statistically significant based on the results of the t-test that was carried out. Differences in text comprehension results based on university were obtained using the Welch One-Way ANOVA test, presented in Table 7 below.

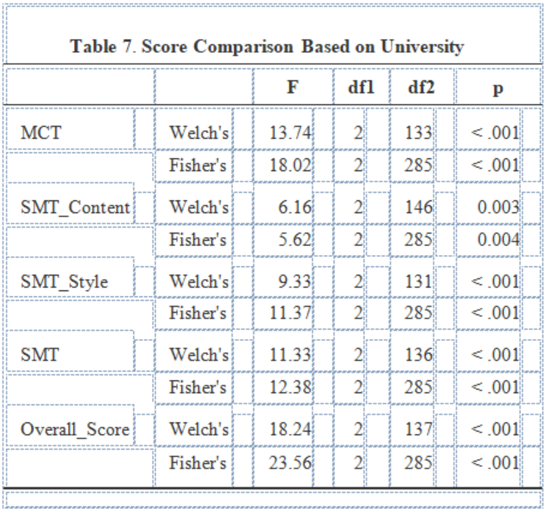


Table 7. Score Comparison Based on University

		F	df1	df2	p
MCT	Welch's	13.74	2	133	< .001
	Fisher's	18.02	2	285	< .001
SMT_Content	Welch's	6.16	2	146	0.003
	Fisher's	5.62	2	285	0.004
SMT_Style	Welch's	9.33	2	131	< .001
	Fisher's	11.37	2	285	< .001
SMT	Welch's	11.33	2	136	< .001
	Fisher's	12.38	2	285	< .001
Overall_Score	Welch's	18.24	2	137	< .001
	Fisher's	23.56	2	285	< .001

Table 7 above shows the statistical results of the hypothesis test for significant differences between universities using the Welch One-Way ANOVA test. The Welch statistic was chosen because the homogeneity assumption was not met, but the normality assumption was met.

The Welch One-Way ANOVA test results show significant differences between universities in all observed variables, namely MCT, SMT_Content, SMT_Style, SMT, and Overall_Score (p<0.05). It means there are significant differences in scores between universities in all aspects measured. In addition, the Fisher's test results also show significant differences between universities for all variables. It confirms the findings from the Welch One-Way ANOVA test that there are significant differences in university scores.

Some factors may contribute to these differences, such as differences in teaching methods, curriculum, or students' level of preparation. Therefore, further research is needed to understand the factors that influence these differences and university strategies to improve student performance in the various aspects measured. Tests of the normality and homogeneity assumptions of the data are presented in Figure 2 and Table 8 below.

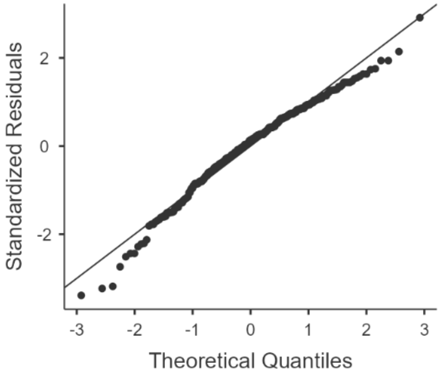


Figure 2: Normality Q-Q Plot of One-Way ANOVA Test

Figure 2 shows that the data is normally distributed. That is depicted by the tending density of the plot approaching a linear line. Additionally, Table 8 presents homogeneity test results from data variance.

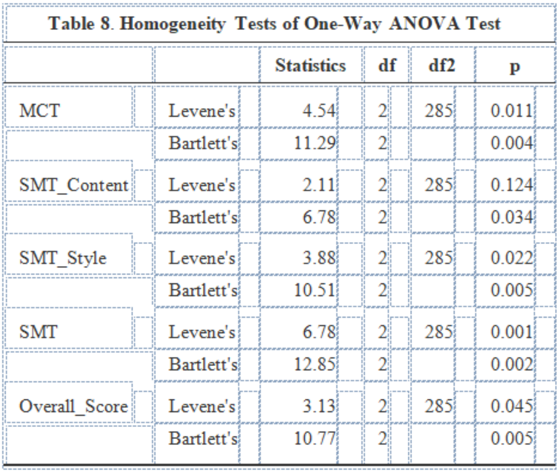


Table 8. Homogeneity Tests of One-Way ANOVA Test

		Statistics	df	df2	p
MCT	Levene's	4.54	2	285	0.011
	Bartlett's	11.29	2		0.004
SMT_Content	Levene's	2.11	2	285	0.124
	Bartlett's	6.78	2		0.034
SMT_Style	Levene's	3.88	2	285	0.022
	Bartlett's	10.51	2		0.005
SMT	Levene's	6.78	2	285	0.001
	Bartlett's	12.85	2		0.002
Overall_Score	Levene's	3.13	2	285	0.045
	Bartlett's	10.77	2		0.005

The results of the homogeneity test using Levene's test for several variables analyzed have been analyzed in the table. The homogeneity assumption is that the variability between groups of data is the same or homogeneous. Based on the test results, MCT, SMT_Style, SMT, and Overall_Score do not meet the homogeneity assumption. The results of the Levene and Bartlett tests for these four variables show that the p-value obtained is less than the specified significance level (0.05), so there is strong enough evidence to state that the data in these variables are not homogeneous. On the other hand, for the SMT_Content variable, the Levene and Bartlett test results show a p-value greater than 0.05, so there needs to be more substantial evidence to conclude that the data in the SMT_Content variable is not homogeneous. Therefore, the Welch statistic was chosen because the data variances were assumed to be unequal.

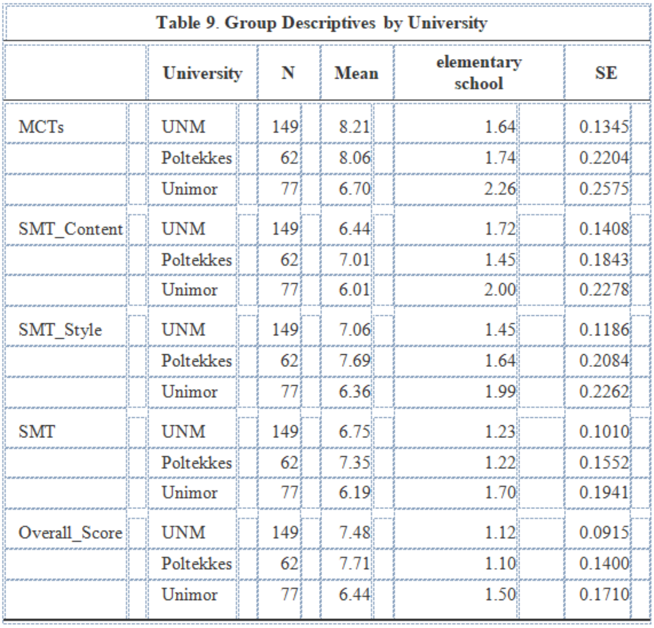


Table 9. Group Descriptives by University

	University	N	Mean	elementary school	SE
MCTs	UNM	149	8.21	1.64	0.1345
	Poltekkes	62	8.06	1.74	0.2204
	Unimor	77	6.70	2.26	0.2575
SMT_Content	UNM	149	6.44	1.72	0.1408
	Poltekkes	62	7.01	1.45	0.1843
	Unimor	77	6.01	2.00	0.2278
SMT_Style	UNM	149	7.06	1.45	0.1186
	Poltekkes	62	7.69	1.64	0.2084
	Unimor	77	6.36	1.99	0.2262
SMT	UNM	149	6.75	1.23	0.1010
	Poltekkes	62	7.35	1.22	0.1552
	Unimor	77	6.19	1.70	0.1941
Overall_Score	UNM	149	7.48	1.12	0.0915
	Poltekkes	62	7.71	1.10	0.1400
	Unimor	77	6.44	1.50	0.1710

Table 9 provides very relevant information about the differences in text comprehension results between three different universities, namely Universitas Negeri Makassar (UNM), Makassar Health Polytechnic (Poltekkes), and Universitas Timor (Unimor) using several evaluation variables.In terms of the Multiple-Choice Test Score (MCT), although the mean difference between UNM and Poltekkes is not significant, UNM has a slightly higher mean score than Unimor, indicating better performance based on this test.

In the Summarizing Test-Content Score (SMT_Content) and Summarizing Test-Style Score (SMT_Style) variables, Poltekkes shows a significantly higher mean value than UNM and Unimor. It indicates that Poltekkes students tend to have better performance in understanding text content and writing style compared to students at other universities. The same can be seen in the Summarizing Test Total Score (SMT) and Overall Score (Overall_Score) variables. Poltekkes again shows a higher mean value than UNM and Unimor. It indicates that the Poltekkes had better overall academic performance in the exams observed.

Based on the data, Poltekkes appears to have better academic performance than UNM and Unimor in various aspects measured, including content understanding, writing style, and overall. These differences reflect differences in teaching methods, curricula, or student characteristics between the universities.

Comparison between students and ChatGPT

Table 10 presents a comparison of the average text comprehension results between students and ChatGPT in several aspects measured, namely MCT (Multiple Choice Test Score), SMT_Content (Summarizing Test-Content Score), SMT_Style (Summarizing Test-Style Score), SMT (Summarizing Test Total Score), and

Overall_Score (Overall Score). From this data, it can be seen that ChatGPT has a higher average score than students in all aspects measured.

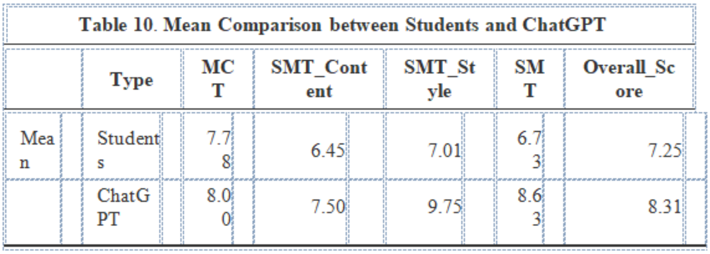


Table 10. Mean Comparison between Students and ChatGPT

	Type	MCT	SMT_Content	SMT_Style	SMT	Overall_Score
Mean	Students	7.78	6.45	7.01	6.73	7.25
	ChatGPT	8.00	7.50	9.75	8.63	8.31

Based on MCT results, ChatGPT has an average score of 8.00. Meanwhile, the student only reaches an average score of 7.78. Something similar also happens with SMT_Content ChatGPT's average score of 7.50, while students only reach an average score of 6.45. Significant improvements are also seen in SMT_Style, ChatGPT's average score of 9.75. Meanwhile, students only reach an average score of 7.01. Furthermore, in SMT (Summarizing Test Total Score), ChatGPT is also superior, with an average score of 8.63. Meanwhile, the student reached an average score of 6.73. Even on Overall_Score, ChatGPT owns an average score of 8.31, while students only reach an average score of 7.25. That thing is more carry on shown in Figure 3 below.

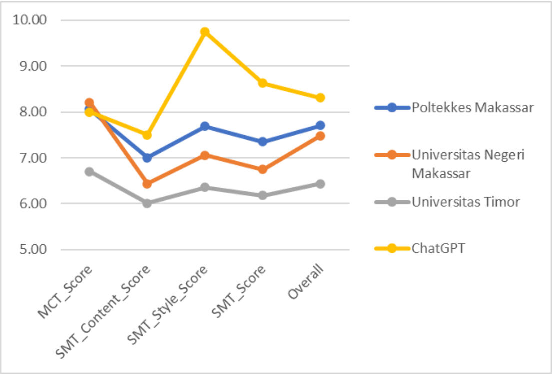
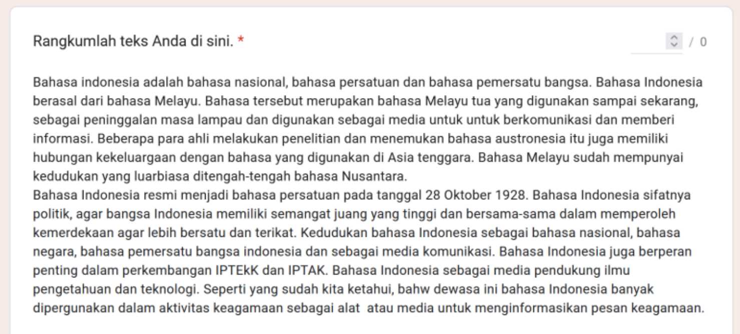


Figure 3: **Level of Text Comprehension by Students and ChatGPT**

Based on the matter, ChatGPT is significantly better in process return text than student based on the aspect being measured. Figure 3 shows that ChatGPT outperforms almost all aspect assessments. However, on aspects, MCT_Score means UNM students are a few times taller than ChatGPT and two other universities. Meanwhile, the Makassar Health Polytechnic has an almost equal position average score ChatGPT. Apart from that, the average of the University of Timor is visibly different significantly among others, with a range score of only 6.00 – 7.00.

It shows that ChatGPT can give summary text with more OK, understand more text, and produce more appropriate answers compared to the student. However, although

ChatGPT is superior in this matter, the presence of man in the world of education Language is still significant Because man's ability to interpret and apply information in more context areas that are not can entirely be replaced by AI technology such as ChatGPT. Figure 4 and Figure 5 show the results summary of students and ChatGPT.

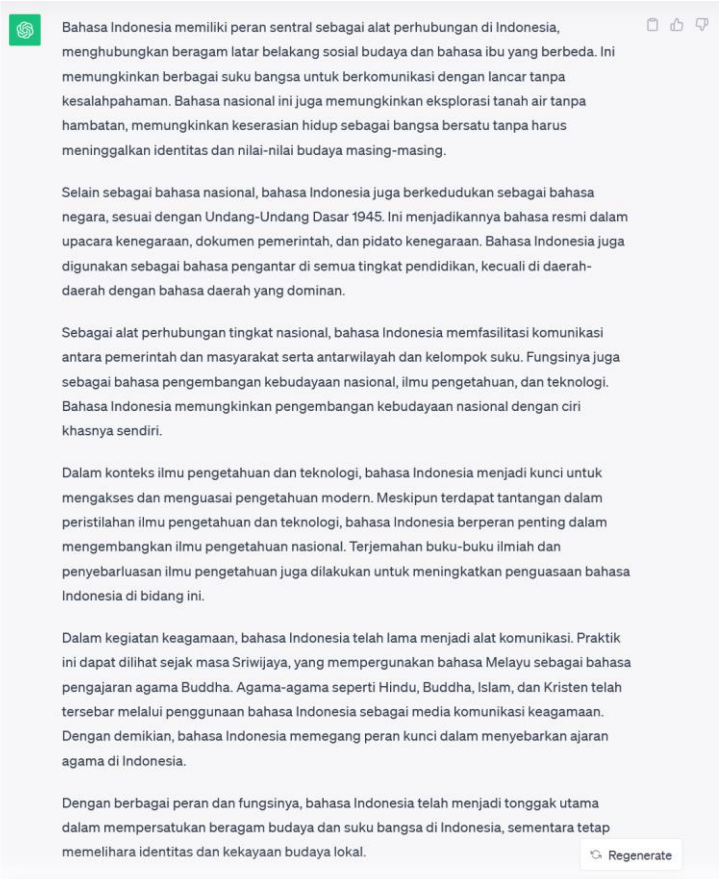


Rangkumlah teks Anda di sini. *

Bahasa indonesia adalah bahasa nasional, bahasa persatuan dan bahasa pemersatu bangsa. Bahasa Indonesia berasal dari bahasa Melayu. Bahasa tersebut merupakan bahasa Melayu tua yang digunakan sampai sekarang, sebagai peninggalan masa lampau dan digunakan sebagai media untuk untuk berkomunikasi dan memberi informasi. Beberapa para ahli melakukan penelitian dan menemukan bahasa austronesia itu juga memiliki hubungan kekeluargaan dengan bahasa yang digunakan di Asia tenggara. Bahasa Melayu sudah mempunyai kedudukan yang luarbiasa ditengah-tengah bahasa Nusantara.

Bahasa Indonesia resmi menjadi bahasa persatuan pada tanggal 28 Oktober 1928. Bahasa Indonesia sifatnya politik, agar bangsa Indonesia memiliki semangat juang yang tinggi dan bersama-sama dalam memperoleh kemerdekaan agar lebih bersatu dan terikat. Kedudukan bahasa Indonesia sebagai bahasa nasional, bahasa negara, bahasa pemersatu bangsa indonesia dan sebagai media komunikasi. Bahasa Indonesia juga berperan penting dalam perkembangan IPTEkK dan IPTAK. Bahasa Indonesia sebagai media pendukung ilmu pengetahuan dan teknologi. Seperti yang sudah kita ketahui, bahw dewasa ini bahasa Indonesia banyak dipergunakan dalam aktivitas keagamaan sebagai alat atau media untuk menginformasikan pesan keagamaan.

Figure 4: **Summary Result by Student No. 195 (Best Score)**



Bahasa Indonesia memiliki peran sentral sebagai alat perhubungan di Indonesia, menghubungkan beragam latar belakang sosial budaya dan bahasa ibu yang berbeda. Ini memungkinkan berbagai suku bangsa untuk berkomunikasi dengan lancar tanpa kesalahpahaman. Bahasa nasional ini juga memungkinkan eksplorasi tanah air tanpa hambatan, memungkinkan keserasian hidup sebagai bangsa bersatu tanpa harus meninggalkan identitas dan nilai-nilai budaya masing-masing.

Selain sebagai bahasa nasional, bahasa Indonesia juga berkedudukan sebagai bahasa negara, sesuai dengan Undang-Undang Dasar 1945. Ini menjadikannya bahasa resmi dalam upacara kenegaraan, dokumen pemerintah, dan pidato kenegaraan. Bahasa Indonesia juga digunakan sebagai bahasa pengantar di semua tingkat pendidikan, kecuali di daerah-daerah dengan bahasa daerah yang dominan.

Sebagai alat perhubungan tingkat nasional, bahasa Indonesia memfasilitasi komunikasi antara pemerintah dan masyarakat serta antarwilayah dan kelompok suku. Fungsinya juga sebagai bahasa pengembangan kebudayaan nasional, ilmu pengetahuan, dan teknologi. Bahasa Indonesia memungkinkan pengembangan kebudayaan nasional dengan ciri khasnya sendiri.

Dalam konteks ilmu pengetahuan dan teknologi, bahasa Indonesia menjadi kunci untuk mengakses dan menguasai pengetahuan modern. Meskipun terdapat tantangan dalam peristilahan ilmu pengetahuan dan teknologi, bahasa Indonesia berperan penting dalam mengembangkan ilmu pengetahuan nasional. Terjemahan buku-buku ilmiah dan penyebarluasan ilmu pengetahuan juga dilakukan untuk meningkatkan penguasaan bahasa Indonesia di bidang ini.

Dalam kegiatan keagamaan, bahasa Indonesia telah lama menjadi alat komunikasi. Praktik ini dapat dilihat sejak masa Sriwijaya, yang mempergunakan bahasa Melayu sebagai bahasa pengajaran agama Buddha. Agama-agama seperti Hindu, Buddha, Islam, dan Kristen telah tersebar melalui penggunaan bahasa Indonesia sebagai media komunikasi keagamaan. Dengan demikian, bahasa Indonesia memegang peran kunci dalam menyebarkan ajaran agama di Indonesia.

Dengan berbagai peran dan fungsinya, bahasa Indonesia telah menjadi tonggak utama dalam mempersatukan beragam budaya dan suku bangsa di Indonesia, sementara tetap memelihara identitas dan kekayaan budaya lokal.

Figure 5: **Summary Results by ChatGPT**

Discussion

The research results show that group men and women understand text similarly based on existing data. The difference between men and women stands out the most in SMT_Content and SMT_Style, with women having a slightly higher average. However, this difference is not statistically significant based on the results of the t-test that was carried out. Furthermore, noteworthy score distinctions emerged among universities across all the measured dimensions. Moreover, supplementary research reveals that humans exhibit a preference for incorporating diverse sentence structures as opposed to ChatGPT. While the average sentence length does not emerge as a distinguishing factor between the two groups, notable distinctions are evident in the standard deviation of sentence length within a specific paragraph, as well as in the median difference (in terms of word count) between a particular sentence and its subsequent counterpart (Desaire et al., 2023). AI literacy is becoming increasingly important, both for the general public and professionals, so that they make better decisions about using AI, avoid misuse of AI, and contribute to the development of this technology. Individual factors reflect AI competencies: technical understanding, critical judgment, and practical application (Laupichler et al., 2023, p. 1). Findings from semi-structured interviews with a dozen instructors at a higher education institution in Hong Kong highlight several vital insights. These insights underscore the significance of educators' proficiency and confidence in employing AI-based teaching tools, shed light on the obstacles and anxieties confronted by language instructors, and underscore the demand for specialized and personalized support in this context (Kohnke et al., 2023, p. 1).

Additionally, the study found that ChatGPT was significantly better at reprocessing text than students based on the measured aspects. ChatGPT outperforms almost every aspect of the assessment. However, in the MCT_Score aspect, the average for UNM students is slightly higher than ChatGPT and the other two universities. Meanwhile, the Makassar Health Polytechnic almost matches the average ChatGPT score. Apart from that, the University of Timor's average appears to be significantly different from the others with a score range of only 6.00 – 7.00. This finding is in line with research, which found that ChatGPT summaries obtained the best scores in terms of content and writing style, with scores respectively 3 and 2.5 points higher than student scores (Vázquez-Cano et al., 2023). BingChat, Bard, and ChatGPT outperform Vietnamese students in English language proficiency (Dao, 2023). The development of AI literacy helps reduce the digital gap between individuals and groups with access to and understanding of AI and those without (S.-C. Kong et al., 2022, p. 1). This approach enables a broader spectrum of individuals to engage in AI development and utilization, fostering inclusivity within the AI era. Utilizing social robots as educational companions has been substantiated as an effective method for enhancing students' comprehension of AI principles (Su & Zhong, 2022, p. 1). The University of Florida (UF) is actively integrating AI into its curriculum and creating avenues for student involvement in designated areas of AI literacy, irrespective of their academic discipline (Southworth et al., 2023, p. 1).

Conclusion

Based on existing data, there is no significant difference between male and female groups in text comprehension. However, the differences between men and women appear to be most pronounced in SMT_Content and SMT_Style, with women having slightly higher averages. In addition, ChatGPT was significantly better at reprocessing text than students based on the measured aspects. ChatGPT outperforms almost every aspect of the assessment.

This research contributes to developing the Indonesian curriculum by utilizing technology such as Artificial Intelligence (AI). This research can help assess the extent to which the education curriculum in Indonesia has succeeded in developing tertiary students' text comprehension abilities. By comparing students' text comprehension abilities with ChatGPT, this research can provide insight into whether the existing curriculum is adequate in developing these abilities or needs changes. The results of this research can influence education policy in Indonesia. The government can use these findings as a basis for making better policies to improve the quality of education.

This research implies that Indonesian students need to gain more understanding of texts compared to ChatGPT. The first implication is the need to revise and develop the educational curriculum. It may include improving course curricula focusing on text comprehension, using more practical learning methods, or integrating technology such as ChatGPT into the learning process. Therefore, future research can examine text comprehension abilities in more specific contexts, such as scientific texts, journalism, literature, or specific scientific disciplines. It can provide more detailed insight into students' strategies for overcoming difficulties in understanding texts. In addition, future research will be conducted on the broader impact of the use of Artificial Intelligence technology in language education on the development of student text comprehension and the potential social and ethical impacts.

Acknowledgments

The author would like to thank the Chancellor of Makassar State University, the Dean of the Faculty of Languages and Literature, and the Chair of the Institute for Research and Community Service for permission so that this research can be carried out.

Bibliographic references

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy, 19(8), 1236-1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Baklazhenko, Y., & Kornieva, Z. (2023). Higher education in the crisis period: A comparative analysis of the Ukrainian experience of online or blended tefl during the pandemic and the war. XLinguae, 16(2), 100-114. https://doi.org/10.18355/xl.2023.16.02.08

Carolus, A., Augustin, Y., Markus, A., & Wienrich, C. (2023). Digital interaction literacy model – Conceptualizing competencies for literate interactions with voice-based AI systems. Computers and Education: Artificial Intelligence, 4 (1), 1-9. https://doi.org/10.1016/j.caeai.2022.100114

Celik, I. (2023). Exploring the Determinants of Artificial Intelligence (AI) Literacy: Digital Divide, Computational Thinking, Cognitive Absorption. Telematics and Informatics, 83 (2003), 1-11. https://doi.org/10.2139/ssrn.4244763

Choi, E. P. H., Lee, J. J., Ho, M. H., Kwok, J. Y. Y., & Lok, K. Y. W. (2023). Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. Nurse Education Today, 125(8). https://doi.org/10.1016/j.nedt.2023.105796

Dai, Y., Liu, A., & Lim, C. P. (2023). Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. Procedia CIRP, 119, 84–90. https://doi.org/10.35542/osf.io/nwqju

Dalalah, D., & Dalalah, O. M. A. (2023). The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. The International Journal of Management Education, 21(2), 100822. https://doi.org/10.1016/j.ijme.2023.100822

Dao, X.-Q. (2023). Performance Comparison of Large Language Models on VNHSGE English Dataset: OpenAI ChatGPT, Microsoft Bing Chat, and Google Bard. ArXiv, 1-12. https://doi.org/10.48550/arXiv.2307.02288

Desaire, H., Chua, A. E., Isom, M., Jarosova, R., & Hua, D. (2023). Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools. Cell Reports Physical Science, 4(6), 1-11. https://doi.org/10.1016/j.xcrp.2023.101426

Duenas, G., Jimenez, S., & Mateus Ferro, G. (2023). You've Got a Friend in ... a Language Model? A Comparison of Explanations of Multiple-Choice Items of Reading Comprehension between ChatGPT and Humans. Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), Bea, 372-381. https://doi.org/10.18653/v1/2023.bea-1.30

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. Computers and Education: Artificial Intelligence, 3(2) 1-7. https://doi.org/10.1016/j.caeai.2022.100055

Giannos, P., & Delardas, O. (2023). Performance of ChatGPT on UK Standardized Admission Tests: Insights From the BMAT, TMUA, LNAT, and TSA Examinations. JMIR Medical Education, 9, e47737, 1-7. https://doi.org/10.2196/47737

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., … Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. Internet of Things and Cyber-Physical Systems, 4, 19-23.

González-Lloret, M. (2023). The road System travelled: Five decades of technology in language education. System, 118. https://doi.org/10.1016/j.iotcps.2023.06.002

Greitemeyer, T., & Kastenmüller, A. (2023). HEXACO, the Dark Triad, and Chat GPT: Who is willing to commit academic cheating? Heliyon, 9(9), 1-9. https://doi.org/10.2139/ssrn.4401953

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. Computers and Education: Artificial Intelligence, 5 ,1-12. https://doi.org/10.1016/j.caeai.2023.100165

Kasneci, E., Sessler, K., Küchemann, S., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences 103(13), 1-13. I https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. Computers and Education: Artificial Intelligence, 5, 1-8. https://doi.org/10.1016/j.caeai.2023.100156

Kong, S.-C., Cheung, W. M.-Y., & Zhang, G. (2022). Evaluating artificial intelligence literacy courses for fostering conceptual learning, literacy and empowerment in university students: Refocusing to conceptual building. Computers in Human Behavior Reports, 7, (2), 1-12. https://doi.org/10.1016/j.chbr.2022.100223

Kong, Z. Y., Adi, V. S. K., Segovia-Hernández, J. G., & Sunarso, J. (2023). Complimentary role of large language models in educating undergraduate design of distillation column: Methodology development. Digital Chemical Engineering, 9, 1-12. https://doi.org/10.1016/j.dche.2023.100126

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the “scale for the assessment of non-experts’ AI literacy” – An exploratory factor analysis. Computers in Human Behavior Reports, 20, 1-10. https://doi.org/10.1016/j.chbr.2023.100338

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and Education: Artificial Intelligence, 3, 1-15. https://doi.org/10.1016/j.caeai.2022.100101

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. International Journal of Management Education, 21(2), 1-13. https://doi.org/10.1016/j.ijme.2023.100790

Lukianenko, V., & Vadaska, S. (2020). Evaluating the Efficiency of Online English Course for First-Year Engineering Students. Revista Romaneasca Pentru Educatie Multidimensionala, 12(2), 62-69. https://doi.org/10.18662/rrem/12.2sup1/290

Mahyudi, A. (2023). Efektivitas Penggunaan Teknologi Dalam Pembelajaran Bahasa Indonesia. ARMADA : Jurnal Penelitian Multidisiplin, 1(2), 122-127. https://doi.org/10.55681/armada.v1i2.393

Mertala, P., Fagerlund, J., & Calderon, O. (2022). Finnish 5th and 6th grade students’ pre-instructional conceptions of artificial intelligence (AI) and their implications for AI literacy education. Computers and Education: Artificial Intelligence, 3, 1-11. https://doi.org/10.1016/j.caeai.2022.100095

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. Research Methods in Applied Linguistics, 2(2), 1-13. https://doi.org/10.2139/ssrn.4373111

Ng, D. T. K., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. Computers and Education: Artificial Intelligence, 3, 1-14. https://doi.org/10.1016/j.caeai.2022.100054

Pramuki, E. (2018). Sejarah Perkembangan Bahasa Indonesia. In Bahasa Indonesia. Universitas Terbuka.

Prastiwi, C. H. W., & Pujiawati, N. (2019). Penggabungan Artificial Intelligence dan Kecerdasan Alami dalam Pembelajaran Keterampilan Menulis Bahasa Inggris. Prosiding Seminar Nasional Pascasarjana, 172-178.

Ramadevi, J., Sushama, C., Balaji, K., Talasila, V., Sindhwani, N., & Mukti. (2023). AI enabled value-oriented collaborative learning: Centre for innovative education. The Journal of High Technology Management Research, 34(2), 23-36. https://doi.org/10.1016/j.hitech.2023.100478

Southworth, J., Migliaccio, K., Glover, J., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and Education: Artificial Intelligence, 4, 1-10. https://doi.org/10.1016/j.caeai.2023.100127

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. Computers and Education: Artificial Intelligence, 4, 1-66. https://doi.org/10.1016/j.caeai.2023.100124

Su, J., & Zhong, Y. (2022). Artificial Intelligence (AI) in early childhood education: Curriculum design and future directions. Computers and Education: Artificial Intelligence, 3 , 1-12. https://doi.org/10.1016/j.caeai.2022.100072

Sweeney, S. (2023). Who wrote this? Essay mills and assessment – Considerations regarding contract cheating and AI in higher education. The International Journal of Management Education, 21(2), 1-7. https://doi.org/10.1016/j.ijme.2023.100818

Tsai, M.-L., Ong, C. W., & Chen, C.-L. (2023). Exploring the use of large language models (LLMs) in chemical engineering education: Building core course problem models with Chat-GPT. Education for Chemical Engineers, 44, 71-95. https://doi.org/10.1016/j.ece.2023.05.001

Vázquez-Cano, E., Ramírez-Hurtado, J. M., Sáez-López, J. M., & López-Meneses, E. (2023). ChatGPT: The brightest student in the class. Thinking Skills and Creativity, 49, 1-12. https://doi.org/10.1016/j.tsc.2023.101380

Wang, X., Li, L., Tan, S. C., Yang, L., & Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. Computers in Human Behavior, 146, 1-11. https://doi.org/10.1016/j.chb.2023.107798

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. Computers & Education, 194, 1-17. https://doi.org/10.1016/j.compedu.2022.104703

Xiao, C., Xu, S. X., Zhang, K., Wang, Y., & Xia, L. (2023). Evaluating Reading Comprehension Exercises Generated by LLMs: A Showcase of ChatGPT in Education Applications. Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), Bea, 610-625. https://doi.org/10.18653/v1/2023.bea-1.52

Yudono, K. D. A. (2023). Keterbatasan Cerita Pendek Horor Karya Artificial Intelligence (AI) pada Perangkat Lunak ChatGPT. DIDAKTIS: Jurnal Pendidikan Bahasa Dan Sastra Indonesia, 1(2), 51-56. https://doi.org/10.33096/didaktis.v1i2.306

Words: 7880
Characters: 54 338 (30,2 standard pages)

Assoc. Prof. Juanda
Indonesian Language and Literature Department
Faculty of Languages and Literature
Universitas Negeri Makassar
South Sulawesi
Indonesia
ORCID ID: https://orcid.org/0000-0003-2058-3314
juanda@unm.ac.id

Iswan Afandi
Indonesian Language and Literature Education Department,
Faculty of Science Education
Universitas Timor
East Nusa Tenggara
Indonesia.
ORCID ID: https://orcid.org/0009-0006-0961-4828
iswan@unimor.ac.id