

THE UTILIZATION OF *GENERATIVE ARTIFICIAL INTELLIGENCE* IN THE PREPARATION OF ASSESSMENT QUESTIONS AND THE DEVELOPMENT OF LEARNING MATERIALS IN EDUCATION UNITS

Armeyda Mauli Abdha Ghibran ^{a*)}, Budiyo ^{a)}

^{a)} Universitas Negeri Semarang, Jawa Tengah, Indonesia

^{*)}Corresponding Author: abdhaaghibran@students.unnes.ac.id

Article history: received 19 May 2026; revised 26 June 2026; accepted 09 July 2026

DOI: <https://doi.org/10.33751/jhss.v10i2.212>

Abstract. This research is motivated by the increasing use of Generative AI in the field of education, but the relationship between the use of this technology and teachers' ability to support the learning process still needs to be understood. This research aims to determine the level of utilization of Generative AI and the relationship between factors related to the use of Generative AI in the preparation of assessment questions and the development of teaching materials in junior high schools, describe the forms of utilization of Generative AI in the preparation of assessment questions and the development of teaching materials; as well as identify the challenges, obstacles, and needs of teachers in utilizing this technology. This study uses a quantitative method with a descriptive correlational design. The research sample consisted of 62 teachers. Data were collected thru questionnaires and analyzed using descriptive statistics and correlation tests with the help of SPSS version 25. The research results show a positive and significant relationship between the use of Generative AI and teachers' ability to develop assessment questions. These findings indicate that Generative AI can function as a supportive technology to assist teachers in creating assessment questions more effectively and efficiently. Additionally, Generative AI can be utilized as an alternative tool to develop various assessment items, adjust the difficulty level of questions, and support the evaluation process. Therefore, educational institutions can consider integrating Generative AI as a supportive resource to enhance the quality of assessment development. However, schools must ensure that the use of Generative AI is implemented wisely and remains aligned with learning objectives, content relevance, and student characteristics. These findings indicate that Generative AI has been widely used by teachers and is related to various aspects that support the effectiveness of assessment preparation and learning material development.

Keywords: *Generative AI*, Assessment, Learning Materials

I. INTRODUCTION

The rapid development of information and communication technology has had an essential impact on educational practices in various countries, including Indonesia. The integration of digital technology in learning is one of the important indicators in improving the quality of education and developing competencies in the 21st century. In recent years, Artificial Intelligence (AI) has developed very rapidly and has begun to be widely used in various fields, including education. One of the parts of AI that has received great attention is Generative Artificial Intelligence, which is a technology that is able to create new content in the form of text, evaluation questions, and learning materials automatically based on user commands or instructions.[1][2], [3]

Generative Artificial Intelligence offers new opportunities for teachers to support pedagogical tasks, especially in learning planning and evaluation of learning outcomes. This technology can be used to help teachers prepare various competency test questions, according to the indicators of learning achievement and the cognitive level of students. In addition, Generative [4]AI can also be used to develop

learning materials in the form of material summaries, teaching modules, and learning support materials tailored to student needs.[5]

In the context of junior high school education, the preparation of competency test questions and the development of learning materials are an important part of the teacher's professional duties. Competency test questions function as a tool to measure the achievement of learning objectives, while learning materials play a role as the main means in helping students understand concepts and develop thinking skills. However, teachers often face time constraints, high administrative burdens, and limited resources in compiling quality questions and learning materials.[6]

The success of educational technology integration is greatly influenced by the pedagogic and technological competence of teachers. Teachers play an important role in determining the quality of evaluation questions and learning materials used in the teaching and learning process. In the use of Generative [7]AI, teachers are expected not only to master the use of technology as a tool, but must have adequate pedagogical understanding so that the questions and materials

produced remain valid, relevant, and in accordance with the characteristics of students.[5], [8]

Despite its great potential, the adoption rate of Generative AI by school teachers is still relatively low. The study revealed that many teachers do not have adequate AI literacy, especially related to the use of generative AI in the context of evaluation and development of teaching materials. In addition, it highlights teachers' concerns regarding the accuracy of content, the potential for plagiarism, and ethical aspects in the use of Generative [9][4]AI in the educational environment. Therefore, clear guidelines and policies are needed so that the use of this technology can be carried out responsibly.[10]

These findings show that there is a gap between the potential of Generative AI in helping teachers prepare competency test questions and learning materials and the real practice of using them by junior high school teachers. This gap points to the need for a more in-depth empirical study of how teachers understand, utilize, and deal with the challenges of using Generative AI in their pedagogical tasks.

Based on initial research through literature studies and limited observations, it was found that most junior high school teachers still prepare competency test questions and learning materials manually or using basic technology, such as word processing and presentation applications. In the Semarang Regency education unit, teachers are also still compiling questions manually so that the process is not practical and efficient. The use of Generative AI to produce question banks, evaluation grids, and draft learning materials is still very limited and has not been systematically integrated in learning planning.

This research was carried out to obtain an analysis related to the level of teachers' understanding in the education unit of the concept and use of Generative AI in the preparation of competency test questions and the development of learning materials, describe the form of its use in compiling questions and developing learning materials, as well as identify obstacles, challenges, and needs of teachers in utilizing the technology. In particular, this study seeks to answer questions about the level of teachers' understanding of Generative AI, the form of its use in the preparation of competency test questions and the development of learning materials, as well as the pedagogical, technical, and ethical obstacles faced in its application.

The results of this study are intended to contribute to the development of the study of Information Technology Education, which focuses on the use of Generative AI in the context of learning assessment and teaching material development. In addition, the results of this study are expected to provide recommendations for teachers and schools in utilizing Generative AI effectively and ethically to prepare competency test questions and learning materials, as well as become a basis for policy makers in formulating training programs and guidelines for the use of Generative AI in schools.

II. RESEARCH METHOD

This study uses a quantitative approach with a type of correlational descriptive research. Quantitative research is

used in the form of testing the relationship between variables through numerical data collection and statistical analysis. This study aims to analyze the relationship between the use of Generative Artificial Intelligence (Generative AI) by teachers and the ability to prepare assessment test questions and develop learning materials. The research was carried out in the education unit, Semarang Regency, from March to May 2026. The research respondents were teachers in educational units who were involved in the preparation of assessment test questions and the development of learning materials. The research data consisted of primary data obtained directly from respondents through the distribution of questionnaires regarding the use of the Generative AI platform in the preparation of assessment questions and the development of learning materials.[11]

The main data collection instrument is in the form of a closed questionnaire which is compiled based on a grid of variables for the use of Generative AI and the ability to compile competency test questions and learning materials. According to , the questionnaire is a data collection tool by providing a list of questions or statements filled in by respondents. To complete the data, this study also used semi-structured interviews. The interview was conducted to obtain information about teachers' understanding, experience of using Generative AI, and the obstacles faced. Observation was used to observe the practice of using Generative AI in the preparation of questions and learning materials, while documentation studies were carried out by analyzing the resulting learning tools.Sugiyono (2013)

The research instruments used have passed validity and reliability testing. The validity of the items on the questionnaire was analyzed using *Pearson's Product Moment* correlation by comparing the *value of r* calculated against *r* of the table at a significance level of 5%, namely *r* calculated $> r$ of the table (0.444). Meanwhile, the reliability of the instrument was tested using the Cronbach's Alpha coefficient, with the reliability criterion if the Cronbach's Alpha value is more than 0.70.

Data were analyzed using IBM SPSS Statistics version 25. The initial stage of analysis is carried out through descriptive statistics which aims to describe the characteristics of the data using mean values, standard deviation, frequency and percentage as indicators. In the next stage, a *Kolmogorov-Smirnov* normality test is carried out as a prerequisite for further analysis. The results of the normally distributed data were analyzed by the parametric test, namely the *Pearson Product Moment* correlation test, while the results that showed the data were not normally distributed were analyzed using a non-parametric test, namely the *Spearman Rank test* (Spearman Correlation test). Hypothesis H_0 is rejected if the significance value (*p*) is less than 0.05 which means that there is a significant relationship between the two research variables.

III. RESULT AND DISCUSSION

In this study, it began by conducting a questionnaire trial in different classes. The test of the measuring tool aims to test the validity and reliability of the questionnaire so that the measurement results obtained can be accounted for. After all

the data is collected, the next stage is to carry out scoring to conduct further data analysis. Data analysis is carried out through several stages, namely re-checking the data that has been collected, scoring respondents' answers, data input and processing in accordance with the purpose of analysis, testing instruments using SPSS version 25 and equipped with interpretation of the data that has been obtained.

[13]) stated that in the data analysis method there is a need for several things to be done, including instrument reliability and validity test techniques, data normality test techniques, hypothesis tests and correlation techniques. Data qualification and data analysis requirements test is carried out by converting data into numerical form. Before the hypothesis test is carried out, the analysis requirements test is first carried out. In the use of parametric statistics, one of the conditions that must be met is that the data obtained must be distributed normally.

Results of Validity Test Analysis

Testing the validity of the questionnaire using generative AI and the ability to compile assessment questions using product moment correlation techniques.

Table I
Test the Validity of the Use of Generative AI

No	r hitung	r tabel	Kesimpulan
1.	0,741	0,444	Valid
2.	0,673	0,444	Valid
3.	0,658	0,444	Valid
4.	0,672	0,444	Valid
5.	0,633	0,444	Valid
6.	0,473	0,444	Valid
7.	0,569	0,444	Valid
8.	0,747	0,444	Valid
9.	0,508	0,444	Valid
10.	0,535	0,444	Valid
11.	0,548	0,444	Valid
12.	0,549	0,444	Valid
13.	0,539	0,444	Valid
14.	0,660	0,444	Valid
15.	0,615	0,444	Valid
16.	0,577	0,444	Valid
17.	0,658	0,444	Valid
18.	0,531	0,444	Valid
19.	0,707	0,444	Valid
20.	0,686	0,444	Valid

Based on the validity test of 20 questions of the questionnaire using Generative AI, it was known that the value of the *Pearson Correlation* calculation $\geq$ r table was 0.444, so that all questionnaires on the use of generative AI were declared valid as research questionnaires.

Based on the validity test of 20 questions of the questionnaire using Generative AI, it was known that the value of the *Pearson Correlation* calculation $\geq$ r table was 0.444, so that all questionnaires on the use of generative AI were declared valid as research questionnaires.

Table II
Test the Validity of the Ability to Prepare Assessment

No	Pearson Correlation	Nilai Sig.	Kesimpulan
1.	0,781	0,444	Valid
2.	0,833	0,444	Valid
3.	0,729	0,444	Valid
4.	0,772	0,444	Valid
5.	0,821	0,444	Valid
6.	0,668	0,444	Valid
7.	0,854	0,444	Valid
8.	0,802	0,444	Valid
9.	0,738	0,444	Valid
10.	0,749	0,444	Valid
11.	0,727	0,444	Valid
12.	0,833	0,444	Valid
13.	0,756	0,444	Valid
14.	0,791	0,444	Valid
15.	0,767	0,444	Valid
16.	0,768	0,444	Valid
17.	0,858	0,444	Valid
18.	0,802	0,444	Valid
19.	0,814	0,444	Valid
20.	0,795	0,444	Valid

Questions

Reliability Test Analysis Results

Table III
Questionnaire Reliability Test of Generative AI Use

Reliability Statistics	
Cronbach's Alpha	N of Items
0,913	20

Table IV
Test the Reliability of the Ability to Prepare Assessment Questions

Reliability Statistics	
Cronbach's Alpha	N of Items
0,967	20

Based on the reliability test for the questionnaire using Generative AI, a cronbach alpha value of 0.967 was obtained, while for the reliability test of the questionnaire on the ability to compile competency questions, a value of 0.967 was obtained, so it had a high reliability value.

Characteristics of Respondents by Age

From the results of the data processing, the description of the age of teachers in the education unit, Semarang Regency is as follows:

Table V
Characteristics of Respondents by Age

Remarks	Quantity	Percentage
20-25 years	12	19,4
26-35 years old	25	40,3
36-45 years old	18	29,1
46-55 years old	7	11,2
Quantity	62	100%

Source: Primary Data 2026

Based on table 5 above, it can be seen that most of the respondents are 26-35 years old, namely 25 people or 40.3%, followed by teachers in the age category of 36-45 years old as many as 29.1%, then 20-25 years old as many as 12 people or 19.4%, and the rest are aged 46-55 years as much as 11.2%. Thus, the average respondent is between 26-35 years old.

Characteristics of Respondents Based on Education

Based on education, the description of the respondent's education can be presented in the table as follows:

Table VI
Characteristics of Respondents Based on Education

Remarks	Quantity	Percentage
D3	4	6,5
S1	54	87
S2	4	6,5
S3	0	0
Quantity	62	100%

Source: Primary Data 2026

Based on table 6 above, it is known that most of the respondents are S1 educational, namely 54 people or 87%, followed by S2 and D3 as much as 6.5%.

Characteristics of Respondents Based on Length of Service

Based on education, the description of the respondent's education can be presented in the table as follows:

Table VII
Characteristics of Respondents Based on Length of Service

Remarks	Quantity	Percentage
<10 years	15	24,2
10-20 years	34	54,8
>20 years old	13	20
Quantity	62	100%

Source: Primary Data 2026

Based on table 7 above, it is known that the respondents who have a working period of < 10 years are 15 people or 24.2%, the 10-20 years of service period is 34 people or 54.8%, and the rest have a service period of > 20 years as many as 13 people or 20%. This explains that the average respondent has a working period of 10-20 years.

Characteristics of Respondents Based on Length of Service

Table VIII
Variable Descriptive Test

	N	Min	Max	Red	Std. Deviation
Generative AI	62	54	98	81,23	11,303
Assessment Questions	62	34	96	72,76	18,273
Valid N (listwise)	62				

Based on table 8 above, it can be seen that the descriptive statistical results test shows that the number of research respondents is 62 people. The Generative AI variable obtained a minimum score of 54 and a maximum of 98 with an average score of 81.23 and showed a standard deviation of 11.303. The results show that the level of Generative AI utilization by respondents tends to be high with a relatively homogeneous data distribution. Meanwhile, the Assessment Question variable has a minimum score of 34 and a maximum of 96 with an average of 72.76 and shows a standard deviation of 18.273. This value shows that the respondents' assessment results are in the category of quite good, but have a higher level of variation than the Generative AI variable. Thus, it can be concluded that respondents have a high level of use of Generative AI, while the results of the preparation of assessment questions show a variety of abilities.

Table IX
Normality Test

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Generative AI	,115	62	,042	,950	62	,014
Soal Asesmen	,162	62	,000	,909	62	,000

a. Lilliefors Significance Correction

In accordance with the table of normality test results using *Kolmogorov-Smirnov*, it shows a significant value in the Generative AI variable of 0.042 and the Assessment Question variable of 0.000. In both values, it is smaller than the significance level of 0.05, so it can be concluded that the data on both variables are not normally distributed. Thus, hypothesis testing is then carried out using nonparametric statistical analysis.

Table X
Non-Parametric Correlation Test

Correlations			Generative AI	Soal Asesmen
Spearman's rho	Generative AI	Correlation Coefficient	1,000	,271*
		Sig. (2-tailed)		,033
	Soal Asesmen	N	62	62
		Correlation Coefficient	,271*	1,000
		Sig. (2-tailed)	,033	
		N	62	62

*. Correlation is significant at the 0.05 level (2-tailed).

Based on table 10 above, it can be seen that the results of the Spearman correlation test show a magnitude of the correlation coefficient of 0.271 with a significance value of 0.033 ($p < 0.05$). In these results, it was explained that there was a significant relationship between the use of Generative AI and the preparation of assessment questions. The value of a positive correlation coefficient explains that the relationship between the two variables is unidirectional, that is, the higher the use of Generative AI, the higher the preparation of assessment questions. However, a value at the correlation coefficient of 0.271 indicates that the strength of the relationship is in the low category. This is due to the age factor and digital understanding by teachers. These findings show that the use of Generative AI is related to teachers' ability to prepare assessment questions. Generative AI can be used as a tool to produce a variety of questions, compile assessment indicators, and develop questions according to the desired level of difficulty. Thus, the use of Generative AI can support teachers in the process of preparing assessment questions more effectively and efficiently. With a high acceptance rate, this study also explains that there are challenges, obstacles, and needs for teachers in utilizing Generative AI. The challenges faced by teachers are continuous with the ability to use Generative AI effectively, such as when compiling commands (*prompts*). Obstacles that often arise are concerns about the accuracy of the information generated by Generative AI, so the information needs to be reviewed before use. Therefore, teachers need to carry out continuous training and mentoring activities in order to make optimal use of Generative AI in the learning process.

IV. CONCLUSION

This research provides implications for educational units in Semarang Regency in the use of Generative AI to support learning activities. The results of the study show that there is a positive and significant relationship between the use of Generative AI and the ability to prepare assessment questions. This finding explains that Generative AI can be used as one of the supporting technologies in helping teachers in compiling assessment questions more effectively and efficiently. The use of Generative AI can be an alternative for teachers in developing a variety of questions, adjusting the difficulty level of the questions, and supporting the learning evaluation process. Therefore, education units can consider the use of Generative AI as one of the supporting means in

improving the quality of learning assessment preparation. Although the use of Generative AI is used as a supporting technology in helping teachers to prepare assessment questions, it was also found that there are limitations in teachers' skills in using AI as well as concerns about the accuracy of the information received. In addition, schools need to ensure that the use of Generative AI is carried out wisely and still pays attention to the suitability of the material, learning objectives, and characteristics of students.

REFERENCES

- [1] E. Subkhan, "Sejarah dan Paradigma Teknologi Pendidikan untuk Perubahan Sosial," *Prenada Media Group*, 2016.
- [2] W. Holmes, M. Bialik, and C. Fadel, "Artificial Intelligence in Education: Promises and Implications for Teaching and Learning," *Center for Curriculum Redesign*, 2019.
- [3] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach," *Pearson*, vol. 4th ed., 2021.
- [4] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, and F. Fischer, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learn. Individ. Differ.*, vol. 103, p. 102274, 2023.
- [5] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education," *International Journal of Educational Technology in Higher Education*, vol. 16(39), pp. 1–27, 2019.
- [6] P. A. Ertmer and A. Ottenbreit-Leftwich, "Teacher technology change: How knowledge, confidence, beliefs, and culture intersect," *Journal of Research on Technology in Education*, vol. 42(3), pp. 255–284, 2010.
- [7] M. J. Koehler and P. Mishra, "What is technological pedagogical content knowledge (TPACK)?," *Contemporary Issues in Technology and Teacher Education*, vol. 9(1), pp. 60–70, 2009.
- [8] S. M. Brookhart, "Assessment Literacy in a Better Assessment Future," *Chinese Journal of Applied Linguistics*, vol. 46(2), pp. 162–179, 2023, doi: 10.1515/cjal-2023-0202.
- [9] X. , et al. Zhai, "AI literacy in education: A systematic review," *Computers and Education: Artificial Intelligence*, vol. 2, p. 10015, 2021.
- [10] UNESCO, "Artificial Intelligence and Education: Guidance for Policy-makers," *UNESCO Publishing*, 2022.
- [11] J. W. Creswell, "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches," *Sage Publications*, vol. 4th ed, 2014.
- [12] Sugiyono, *Metode Penelitian Pendidikan*. Bandung: Alfabeta, 2013.
- [13] S. Azwar, *Metode Penelitian*. Yogyakarta: Pustaka Pelajar, 2004.