Journal of Educational Sciences Vol. 10 No. 4 (April, 2026) 910-926

Journal of Educational Sciences

Journal homepage: https://jes.ejournal.unri.ac.id/index.php/JES

P-ISSN 2581-1657

E-ISSN 2581-2203

Analysis of Differences in AI Literacy Levels among Students at SMPN 1 Purwakarta in IT Classes and Regular Classes

Sri Rahayu, Ulva Elviani*

Information Systems and Technology Education, Indonesia University of Education, Bandung, 40154, Indonesia

ARTICLE INFO

Article history:

Received: 28 Feb 2026
Revised: 06 April 2026
Accepted: 08 April 2026
Published online: 15 April 2026

Keywords:

AI literacy,
AI ethics,
Comparative study,
Ex post facto design,
Educational measurement

* *Corresponding author:*

E-mail: ulva@upi.edu

Article Doi:

https://doi.org/10.31258/jes.10.4.p.910-926

A B S T R A C T

Artificial intelligence (AI) is increasingly integrated into learning, encompassing conceptual understanding, practical skills, ethics, and attitudes/motivation/self-efficacy. This study aimed to describe AI literacy levels among junior high school students in IT and regular classes and to analyze differences between both groups at SMPN 1 Purwakarta. A non-experimental quantitative approach with an ex post facto comparative design was employed involving 132 eighth-grade students (66 IT and 66 regular). Instrument quality was examined using Aiken's V for content validity, corrected item–total correlation, and Cronbach's alpha/KR-20 for reliability. Data were analyzed using descriptive statistics, Shapiro–Wilk normality test, and independent t-test or Mann–Whitney U test based on distribution. Results indicated a significant difference only in the AI Ethics indicator, with IT students scoring higher ($M = 4.04$, $SD = 0.52$) than regular students ($M = 3.82$, $SD = 0.64$; $p = 0.027$; $g = 0.38$). Other indicators (Apply/Use, Attitude/ Motivation/ Self-efficacy, total Likert score, and Understand) showed no significant differences ($p > 0.05$). These findings highlight the importance of strengthening AI ethics across classes and improving measurement of conceptual understanding in future research.

1. Introduction

Artificial intelligence (AI) is developing rapidly and becoming increasingly integrated into daily activities, including in the context of learning at school. The integration of technology-supported learning activities has been shown to enhance students' engagement and technological literacy in classroom environments (Wulansari et al., 2026). Educational technology implemented in classroom instruction can also support the development of students' digital competencies and encourage more active participation in technology-based learning environments

(Samiha & Aksara, 2025). In the field of education, the use of AI not only reflects technological advances, but also changes the way we learn through more adaptive learning support and opens up opportunities to strengthen computational thinking and critical thinking skills. AI literacy is beginning to be considered an important competency that must be developed from primary and secondary education, not only in technical aspects, but also in conceptual understanding and critical reflection on technology (Long & Magerko, 2020; Ng et al., 2021). However, the use of AI also carries ethical and social consequences, such as issues of algorithmic bias, privacy, and data security, so students need to have adequate literacy readiness to be able to use technology safely, responsibly, and reflectively (Walter, 2024; Zhang et al., 2024). The issue of algorithmic bias in education systems can affect the fairness of students' learning opportunities and experiences, making ethical aspects crucial in building an understanding of AI (Baker & Hawn, 2022). Thus, AI literacy should not be understood merely as conceptual knowledge, but as the ability to use AI appropriately, consider ethical aspects, and develop attitudes and self-confidence that support its responsible use.

To measure AI literacy skills, various literacy instruments and frameworks have been developed in recent years. These AI literacy instruments and frameworks include understanding AI concepts, usage skills, AI ethics, and aspects of attitude, motivation, or self-efficacy in interactions with AI (Carolus et al., 2023; Markus et al., 2025; Nong et al., 2024). The literature on the application of AI in education also explains that integrating this technology requires preparation in terms of education and sufficient literacy skills from various parties (Zawacki-Richter et al., 2019).

However, the application of AI literacy in the context of junior high schools has its own challenges, mainly due to differences in access, facilities, and learning experiences between classes in the same school. Learning environments that provide greater exposure to digital technologies may influence students' technological awareness and competencies in technology-based learning contexts (Muraina & Adesanya, 2024). In practice, some schools implement technology-based classes (e.g., IT classes) that tend to provide more intensive ICT exposure than regular classes, raising the question of whether these differences in learning contexts are related to differences in students' AI literacy levels.

Empirical studies on AI literacy still focus more on university students or adult groups, both in terms of measurement and learning intervention, while evidence at the junior high school level is relatively limited (Chiu et al., 2024; Israwati Hamsar et al., 2024; P. Kesuma & Amelia Fransen, 2025). In addition, research that directly compares AI literacy between two different learning groups within the same school, such as IT classes and regular classes, is still rarely conducted systematically. This lack of comparative evidence means that schools do not yet have a clear picture of the condition or level of AI literacy among their students, especially when schools want to design more inclusive and equitable learning policies or programs.

Given this gap, this study aims to describe the level of AI literacy among junior high school students and analyze the differences in AI literacy levels between IT classes and regular classes through the indicators of Understand, Apply/Use, AI Ethics, and Attitude/Motivation/Self-efficacy. In the digital era, the development of students' digital literacy skills is increasingly important to support their ability to interact with various technologies used in learning environments (Wardhani & Asyiah, 2025). Technology-based learning media such as digital modules have also been shown to support the development of students' higher-order thinking skills and improve learning effectiveness in technology-enhanced classrooms (Rosiana et al., 2023). Measurements were conducted using instruments adapted from (P. Kesuma & Amelia Fransen, 2025) and adjusted to the context of junior high school students, so that the measurements remained structured and relevant to the AI literacy competencies discussed in the literature. With this approach, the research aims to produce an empirical picture of the differences (or equality) in AI literacy levels between IT classes and regular classes at the junior high school level.

The specific objective of this study is to examine differences in AI literacy levels between IT and regular class students using four indicators, namely Understand, Apply/Use, AI Ethics, and Attitude/Motivation/Self-efficacy. The Understand indicator measures students' conceptual understanding of the basic principles and characteristics of AI through objective tests. The Apply/Use indicator discusses how students can correctly use and utilize artificial intelligence in learning. The AI Ethics indicator measures students' awareness of ethical issues such as privacy, bias, and responsibility in the use of AI. Meanwhile, the Attitude/Motivation/Self-efficacy indicator describes students' self-confidence and readiness to interact with AI technology.

This study is expected to provide an empirical picture of the level of AI literacy in two different classroom contexts in one school. In addition, this study will provide a reflective basis for schools to understand how variations in AI literacy levels arise between IT classes and regular classes. However, this study does not intend to find a causal relationship.

2. Methodology

This study uses a non-experimental quantitative approach with an ex post facto comparative design, which compares two naturally formed classes (IT class and regular class) without experimental treatment. This approach is appropriate when researchers do not manipulate variables or randomize subjects, but wants to test the differences in results between classes based on conditions that have already occurred (after the fact) through observational data. This practice is also common in educational research that uses ex post facto/causal-comparative non-experimental designs to compare two groups in a single measurement phase (Expósito López, 2020; Sharma & Adhikari, 2022). Because the groups were not randomly formed, the differences found are interpreted as differences in scores between existing groups (non-causal) and may still be influenced by unmeasured initial characteristics.

The study was conducted at SMPN 1 Purwakarta on eighth-grade students. The research activities took place from October 2025 (including preparation, licensing, and instrument development). Data collection was carried out through the distribution of questionnaires from December 17 to 24, 2025. Participation consent was obtained through school approval and student consent when filling out the instruments. Respondents filled out the questionnaires independently according to the instructions provided, with the researchers ensuring that the filling procedures were uniform in each class to minimize technical differences in implementation between classes.

This study was conducted after obtaining permission from the school. Student participation was voluntary, and respondents received a brief explanation of the purpose of the study and their right to not answer or stop filling out the questionnaire at any time without consequences. Data were collected anonymously without including personal identities and were presented in aggregate form to maintain confidentiality.

The research sample consisted of 132 students from two IT classes and two regular classes. All IT classes (two classes) participated as samples, while in regular classes (six classes), two classes were selected, namely classes J and K. The selection of regular classes was based on school operational considerations (e.g., schedule equality, availability of time for filling out forms, and school recommendations) to ensure that the data collection procedure was uniform. Thus, the most appropriate sampling technique is described as purposive sampling based on considerations (purposive sampling) of intact classes, because the selection of regular classes was not done randomly. In the school context, IT classes are positioned as learning groups with an emphasis on technology-based learning, while regular classes follow general learning at the same level. Therefore, differences in scores (if any) could potentially be influenced by unmeasured initial factors, such as previous ICT experience, access to devices, or ICT-related academic achievement. This study did not measure baselines/covariates, so the findings are interpreted as differences in scores between classes that have already been formed, not as evidence of a causal relationship. Consequently, the results of the study are stronger in explaining the context of the schools/cohorts studied and caution is needed if they are generalized to a wider population (Etikan et al., 2016).

The research instrument was adapted from (P. Kesuma & Amelia Fransen, 2025) to measure AI literacy, covering four indicators, namely Understand, Apply/Use, AI Ethics, and Attitude/Motivation/Self-efficacy. The Understand indicator was measured using objective items with dichotomous responses (correct = 1, incorrect = 0), so that the U_total score ranged from 0 to 5. Meanwhile, the Apply/Use, AI Ethics, and Attitude/Motivation/Self-efficacy indicators were measured using a 5-point Likert scale. The composite score for each indicator was calculated as the mean of all items (after reverse scoring for negative items), so that the indicator scores ranged from 1 to 5. The total Likert score (AI_Total) is calculated as the mean of the three Likert indicator scores (A_Mean, E_Mean, and S_Mean), because each indicator consists of five items, AI_Total is equivalent to the mean of all 15 items on the Likert scale. Negative items are reverse scored before analysis to maintain consistency in the direction of interpretation (Antoniou & Alghamdi, 2024). The structure of the AI literacy instrument used in this study, including the indicators measured, number of items, response formats, scoring ranges, and composite score calculations, is summarized in Table 1.

Table 1. AI Literacy Instrument Structure and Scoring Scheme

Indicators	Number of Items	Response Format	Score Range	Composite Score
Understand	5	Dichotomous (0-1)	0-5	Number of correct answers (U_Total)
Apply/Use	5	Likert 1-5	1-5	Item mean (A_Mean)
AI Ethics	5	Likert 1-5	1-5	Item mean (E_Mean)
Attitude/Motivation/ Self-efficacy	5	Likert 1-5	1-5	Item mean (S_Mean)
AI_Total	15	Likert 1-5	1-5	Mean A_Mean, E_Mean, S_Mean

Instrument quality testing was conducted in two stages. First, content validity was assessed through expert judgment by two experts using a 1–5 scale validation sheet. Each item was assessed holistically, taking into account its suitability for the research objectives and AI literacy indicators, the clarity of the wording, the appropriateness of the language used in accordance with the characteristics of junior high school students, and the suitability of the instrument for use in research. The content validity index for each item was calculated using Aiken's V, with the general formula $V = \frac{\sum s}{n\,(c-1)}$, where $s = r - l_o$, is the score/rating given by the expert, l_o is the lowest score on the scale, n is the number of experts, and c is the number of scale categories. Because the scale is 1–5 ($l_o = 1; c = 5$) and there are 2 experts, the calculation is equivalent to $V = \frac{r_1+r_2-2}{8}$. In this study, items with V < 0,70 were categorized as needing revision, while V > 0,70 were considered feasible. Items deemed to need improvement were revised based on expert notes (e.g., to reduce ambiguity/multiple interpretations), then the revised versions were consulted again with experts and declared acceptable for use (without re-filling quantitative scores). Second, empirical testing was conducted through item analysis using Corrected Item–Total Correlation (CITC) and internal reliability testing using Cronbach's alpha (Likert scale) and KR-20 (alpha equivalent) for dichotomous items. Internal reliability was assessed to ensure consistency between items in a single indicator (Tavakol & Dennick, 2011).

Data analysis begins with data screening (checking score ranges, coding consistency, and response completeness). If there are missing responses on some items, they are handled differently depending on the purpose of the analysis. In reliability testing, SPSS applies listwise deletion so that cases with one blank item on a particular indicator can be removed from the calculation (for example, on the

Apply/Use indicator). However, for descriptive analysis and difference tests, composite scores per indicator are calculated using the mean of available items, so that indicator scores remain in the range of 1–5. In this way, respondents with one missing response can still be included in the main analysis (Dong & Peng, 2013). Indicator scores are calculated if respondents fill in at least ≥80% of the items on that indicator (on a 5-item scale, at least 4 items must be filled in). Furthermore, to test the difference in scores between the IT and regular classes, the researchers conducted prerequisite tests (e.g., normality and homogeneity of variance). If the assumptions were met, an independent two-sample t-test was used, if not, the nonparametric Mann–Whitney U alternative was used. This pattern of test selection based on data characteristics is commonly used in the practice of analyzing two independent groups, although caution is still needed in interpreting decisions based on initial normality tests (Chicco et al., 2025; Rochon et al., 2012). To clarify the sequence of procedures carried out in this study, the overall research process from problem formulation to data analysis and conclusion is illustrated in Figure 1.

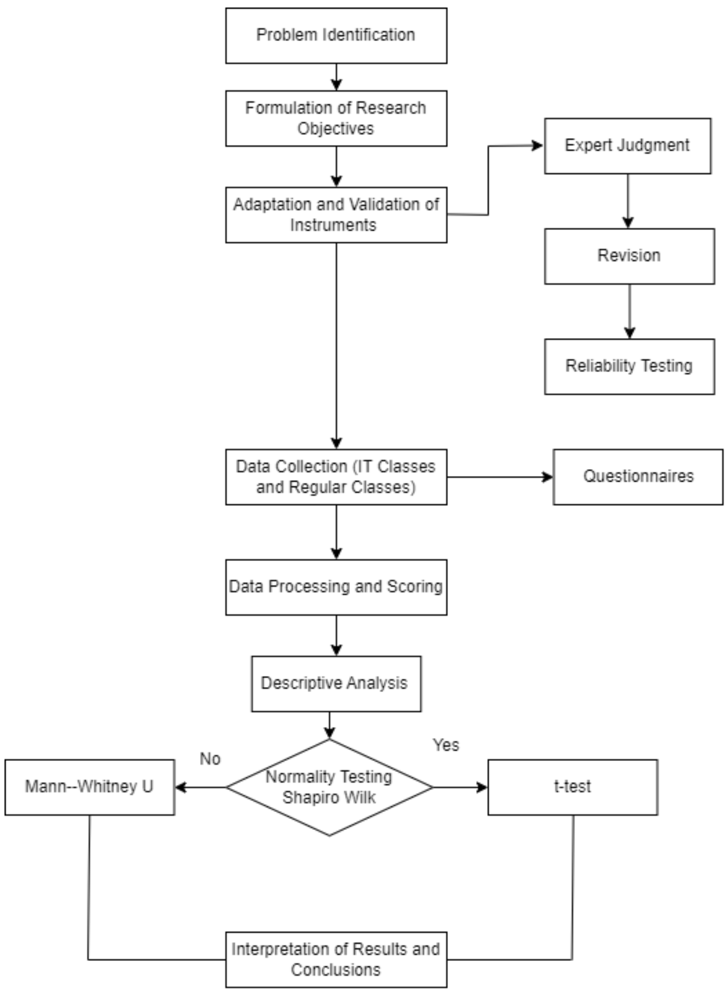

Figure 1. Research Flow

As shown in Figure 1, the research was conducted systematically, beginning with problem formulation, followed by instrument preparation and data collection, and ending with statistical analysis and conclusion drawing.

3. Results and Discussion

The research respondents consisted of 132 eighth-grade students at SMPN 1 Purwakarta who participated in filling out the questionnaire during the period of December 17–24, 2025. The sample was divided based on class type, namely IT classes and regular classes. The IT class consisted of two study groups, and all students were included as respondents. Meanwhile, the regular class was taken from two study groups, namely classes VIII J and VIII K, which were selected from six regular classes at the same level. With this composition, the number of respondents in each class was balanced (IT = 66, regular = 66), so that the comparison of scores between classes could be done proportionally.

Before the main analysis was conducted, the research instrument underwent two stages of testing. The first stage was content validity through expert judgment by two experts using a scale of 1–5, and the content validity index per item was calculated using Aiken's V. Aiken's V values ranged from 0.625 to 0.875 with an average of 0.800. Eighteen of the 20 items met the eligibility criteria, while two items on the Understand indicator were revised according to expert notes to reduce ambiguity/multiple interpretations. The revised version was reconfirmed by experts and declared suitable for use. The second stage was empirical testing using respondent data to assess item quality through corrected item–total correlation (CITC) and internal reliability (Cronbach's alpha/KR-20). To provide an overview of the results of the expert judgment process, the summary of the content validity analysis using Aiken's V for each indicator is presented in Table 2.

Table 2. Summary of Content validity Based on Aiken's V

Indicators	Number of Items	Response Format	Score Range	Composite Score
Understand	5	0,625–0,875	0,775	2
Apply/Use	5	0,750–0,875	0,825	0
AI Ethics	5	0,750–0,875	0,825	0
Attitude/Motivation/Self-efficacy	5	0,750–0,875	0,775	0
Total	**20**	**0,625–0,875**	**0,800**	**2**

Based on Table 2, the content validity index (Aiken's V) per indicator ranges from 0.625 to 0.875 with an overall average of 0.800, indicating that the instrument items are generally considered relevant and suitable for use. Almost all items met the feasibility criteria ($V \geq 0.70$), while two items in the Understand indicator had $V < 0.70$ and were therefore revised following expert notes to reduce ambiguity/multiple interpretations. The revised version was then reconfirmed by experts and declared feasible for use without re-scoring the quantitative scores. After content validity was met, the instrument was then empirically tested using respondent data to ensure item discrimination (CITC) and internal consistency of each indicator (Cronbach's alpha/KR-20). To evaluate the empirical quality of the AI literacy instrument, item discrimination and internal reliability were analyzed using corrected item–total correlation (CITC) and Cronbach's alpha/KR-20. The summary of these empirical test results for each indicator is presented in Table 3.

Table 3. Results of the AI Literacy Instrument Quality Test

Indicators	Number of Items	N (Valid)	Cronbach's Alpha	CITC Range	Brief decision
Understand	5	132	0,283	0,102 - 0,183	Low reliability; interpret with caution
Apply/Use	5	131	0,636	0,311-0,557	Acceptable; all items retained
AI Ethics	5	132	0,674	0,333-0,525	Acceptable; all items retained
Attitude/Motivation/Self-efficacy	5	132	0,644	0,301-0,486	Acceptable; all items retained

Note: N dimensions Apply/Use = 131 because there is 1 missing data. For dichotomous items (0/1) on the Understand indicator, Cronbach's alpha is equivalent to KR-20.

Based on Table 3, item analysis shows that all three dimensions of the Likert scale meet the criteria for corrected total item correlation (CITC ≥0.30). In the Apply/Use dimension (α=0.636; N=131), the CITC value ranged from 0.311 to 0.557, and no item increased the alpha value when deleted (alpha if item deleted 0.498–0.628), so all items were retained. In the AI Ethics (α=0.674; N=132) and Attitude/Motivation/Self-efficacy (α=0.644; N=132), the CITC values were in the range of 0.333–0.525 and 0.301–0.486, respectively. There was no significant increase in alpha when items were deleted (maximum increase of 0.001), so all items were retained for further analysis. In contrast, the Understand indicator showed low reliability (α=0.283) with low CITC values for all items (0.103–0.183), indicating limited item discrimination. Given that the items are dichotomous and there are only five of them, and there is a tendency for the items to be very easy (ceiling effect), the Understand results are reported as exploratory findings and are not used as the main basis for drawing conclusions about differences between classes. Descriptively, the limited response variation was caused by the large number of correct answers on several Understand items, which could affect the low internal reliability.

After instrument testing, descriptive statistics were used to describe students' AI literacy profiles by class (IT vs. regular). Statistics are presented as mean±SD and median (IQR). Likert indicators (E_Mean, A_Mean, S_Mean, AI_Total) range from 1–5, while U_Total represents the number of correct answers (0–5). To provide an overview of the distribution of students' AI literacy scores in each class, the descriptive statistics for each indicator and the total score are presented in Table 4.

Table 4. Descriptive Statistics of Students' AI Literacy Scores Based on Class (IT vs. Regular)

Variable	IT Mean±SD	IT Median (IQR)	Regular Mean±SD	Regular Median (IQR)	Total Mean±SD
E_Mean (AI Ethics)	4,04±0,52	4,00 (1)	3,82±0,64	3,80 (1)	3,93±0,59
A_Mean (Apply/Use)	3,98±0,52	4,00 (1)	4,03±0,56	4,20 (1)	4,00±0,54
S_Mean (Attitude/Motivation/Self-efficacy)	3,89±0,50	3,80 (1)	3,91±0,54	3,80 (1)	3,90±0,52
AI_Total (Total Likert)	3,97±0,38	3,93 (1)	3,92±0,45	3,90 (1)	3,94±0,42
U_Total (Understand; skor 0–5)	4,02±0,81	4,00 (1)	3,86±0,96	4,00 (2)	3,94±0,89

Note: Values are presented as mean±*SD* and median *(IQR)*. IQR = Q3–Q1. E_Mean, A_Mean, S_Mean, and AI_Total are the average scores on the Likert scale (1–5). U_Total is the comprehension test score (number of correct answers; theoretical range 0–5).

Table 4 shows relatively high AI literacy scores, with Likert indicator means close to 4 in both classes. This shows that both IT and regular class students have a positive tendency in the indicators of AI use (Apply/Use), AI ethics, and attitude, motivation, and self-efficacy. The descriptive results also show that there is a significant difference between classes on the AI Ethics indicator (E_Mean), where the IT class has a higher average than the regular class. Meanwhile, on Apply/Use (A_Mean) and Attitude/Motivation/Self-efficacy (S_Mean), the averages of the two classes are relatively the same.

This shows that the level of AI literacy between classes is mainly focused on ethical aspects, rather than on usage or attitudes. In line with this pattern, the AI_Total score in the IT class is slightly higher than in the regular class, but the difference is relatively small. In the Understand (U_Total) indicator, the IT class also had a higher average, although the median of both groups was the same. Thus, the results in the Understand indicator need to be considered by taking into account the dichotomous test score characteristics and the low reliability of the indicator, so that its ability to identify differences between classes is limited. This pattern suggests that differences in AI literacy between the two classes are relatively limited and tend to appear more prominently in the AI ethic indicator.

Before conducting the inter-class difference test, the normality assumption was examined using the Shapiro–Wilk test for each variable in both classes. The results of this normality test were used to determine the appropriate statistical test procedure. Variables that met normality were analyzed using a parametric test (two-sample independent t-test), while variables that did not meet normality were analyzed using a nonparametric alternative (Mann–Whitney U).

Table 5. Results of the Normality Test (Shapiro–Wilk) of AI Literacy Scores Based on Class

Variabel	IT: W	IT: p	Regular: W	Regular: p	Decision (α=0,05)
E_Mean	0,970	0,110	0,974	0,174	Normally distributed (both p>0,05)
A_Mean	0,962	0,043	0,901	<0,001	Not normally distributed
S_Mean	0,971	0,126	0,972	0,147	Normally distributed (both p>0,05)
AI_Total	0,980	0,369	0,986	0,671	Normally distributed (both p>0,05)
U_Total	0,819	<0,001	0,868	<0,001	Not normally distributed

The results presented in Table 5 indicate that E_Mean, S_Mean, and AI_Total meet the normality assumption in both classes ($p > 0.05$). Therefore, differences in these variables between the IT and regular classes were analyzed using an independent two-sample t-test. The selection of the t-test results was based on the homogeneity of variance test (Levene's test), using the equal variances assumed row when the variance was homogeneous and the equal variances not assumed row when the assumption was violated. In contrast, A_Mean did not meet the normality assumption in at least one class, namely the regular class ($p < 0.05$), and was therefore analyzed using the Mann–Whitney U test as a nonparametric alternative. The U_Total variable also did not meet the normality assumption in both classes because it represents a discrete test score (0–5). Consequently, the difference analysis for this variable was also performed using a nonparametric approach. In addition to statistical significance, the analysis also reports effect sizes to provide a practical interpretation of the magnitude of differences between classes (see Tables 6–7).

Table 6. Results of the independent two-sample t-test for E_Mean, S_Mean, and AI_Total scores based on class type

Variable	IT (n=66) Mean±SD	Regular (n=66) Mean±SD	t(df)	p	Mean Diff. (IT–Reg)	95% CI (Lower, Upper)	g (Hedges)
E_Mean (AI Ethics)	4,04±0,52	3,82±0,64	2,234 (130)	0,027	0,227	0,026; 0,429	0,38
S_Mean (Attitude/Motivation/Self-efficacy)	3,89±0,50	3,91±0,54	-0,168 (130)	0,867	-0,015	-0,194; 0,163	-0,04

AI_Total (Total Likert)	3,97±0,38	3,92±0,45	0,753 (130)	0,453	0,055	-0,089; 0,198	0,12

Note: Two-tailed t-test; equal variances assumed row was used because Levene's test showed homogeneous variances (p>0.05). Mean Diff. = mean difference (IT–Regular). g = Hedges' g (effect size).

Based on Table 6, a statistically significant difference was found in the AI Ethics indicator (E_Mean). Students in the IT class showed a higher score (M = 4.04, SD = 0.52) than those in the regular class (M = 3.82, SD = 0.64), $t(130) = 2.234$, $p = 0.027$. The mean difference was 0.227 with a 95% confidence interval [0.026, 0.429]. The effect size was $g = 0.38$, indicating a small to moderate effect. These results suggest that the AI ethics dimension shows the clearest difference between the two classes. Possible factors related to this difference are discussed further in the Discussion section. In contrast, no statistically significant differences were found in the Attitude/Motivation/Self-efficacy indicator (S_Mean), $t(130) = -0.168$, $p = 0.867$, $g = -0.04$, or in the overall AI literacy score (AI_Total), $t(130) = 0.753$, $p = 0.453$, $g = 0.12$. The effect sizes for these variables were very small, indicating that the levels of attitude, motivation, self-efficacy, and overall Likert-based AI literacy scores were relatively similar between IT and regular class students. Based on the results of the normality test, the variables A_Mean (Apply/Use) and U_Total (Understand) did not meet the normality assumption. Therefore, differences between the IT and regular classes were analyzed using the nonparametric Mann–Whitney U test. The results are presented in Table 7, including the median (IQR), U statistic, p-value, and effect size (r).

Table 7. Mann–Whitney U test results based on class (IT vs. regular)

Variabel	IT Median (IQR)	Regular Median (IQR)	U	Z	p	r
A_Mean (Apply/Use)	4,00 (1)	4,20 (1)	2011	-0,767	0,443	0,067
U_Total (Understand; 0–5)	4,00 (1)	4,00 (2)	1998	-0,875	0,381	0,076

Note: Values are presented as median *(IQR)* $r = |Z|/\sqrt{N}$ (N=132).

Based on Table 7, no statistically significant differences were found between the IT class and the regular class in the Apply/Use (A_Mean) and Understand (U_Total) indicators. For the Apply/Use indicator, the median score for the IT class was 4.00 (IQR = 1), while the regular class had a median of 4.20 (IQR = 1). The Mann–Whitney test produced $U = 2011$, $Z = -0.767$, $p = 0.443$, with an effect size of $r = 0.067$, indicating a very small effect. For the Understand indicator, the median score in both classes was 4.00. However, the variation in scores was slightly larger in the regular class (IQR = 2) compared with the IT class (IQR = 1). The test results were also not statistically significant ($U = 1998$, $Z = -0.875$, $p = 0.381$, $r = 0.076$), with a very small effect size. These results indicate that the levels of AI usage and conceptual understanding measured in this study were relatively similar between IT and regular class students. In addition, the interpretation of the Understand indicator should be considered cautiously due to the dichotomous scoring format and the relatively low reliability of the indicator, which may limit its ability to detect differences between groups.

These results indicate that the observed differences represent score variations between existing class groups in an ex post facto comparative design rather than causal effects of instructional treatment. Overall, the results of the difference tests indicate that a statistically significant difference between the IT and regular classes was found only in the AI Ethics indicator. In contrast, the Apply/Use, Attitude/Motivation/Self-efficacy, AI_Total, and Understand indicators did not show statistically significant differences ($p > 0.05$). To provide a clearer comparison of AI literacy scores between classes, a visual summary is presented in Figure 2.

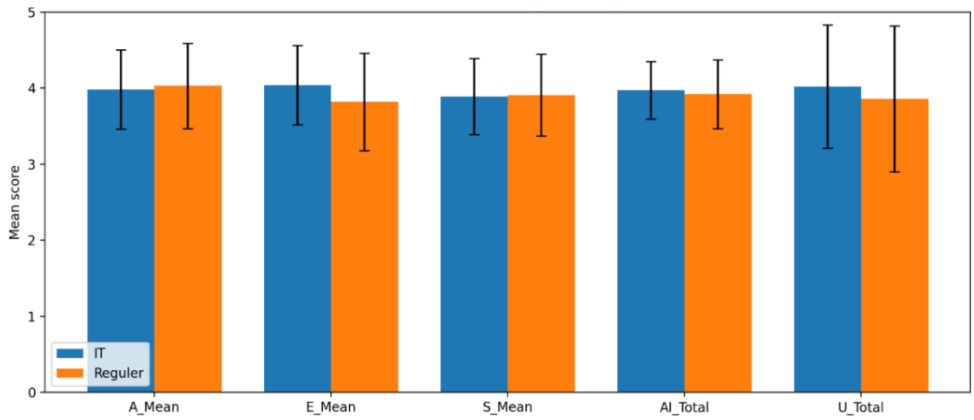

Figure 2. Average Scores of AI Literacy Indicators by Class (IT vs. regular).

Figure 2 illustrates the comparison of the average scores for the five AI literacy indicators between the IT and regular classes. Visually, the largest difference appears in the AI Ethics indicator, where the IT class shows a slightly higher average score than the regular class. This pattern is consistent with the statistical test results, which indicate a significant difference for this indicator. For the other indicators (Apply/Use, Attitude/Motivation/Self-efficacy, AI_Total, and Understand), the average scores between the two classes appear relatively similar. Although the IT class shows a slightly higher mean score on the Understand indicator, the difference is small and should be interpreted cautiously due to the greater variation in responses and the limited reliability of the indicator.

Discussion

This study examined junior high school students' AI literacy across four indicators: Understand, Apply/Use, AI Ethics, and Attitude/Motivation/Self-efficacy, and compare them between IT and regular classes. Descriptively, the three Likert-based indicators show relatively high scores (mean ≥ 3.40 on a 1–5 scale), suggesting that students generally demonstrate a positive tendency toward interacting with AI technologies. This pattern suggests that students already possess a basic readiness to engage with AI in educational contexts. These findings align with the concept of AI literacy as a multidimensional construct that includes not only conceptual knowledge but also usage skills, ethical awareness, and affective aspects such as attitudes, motivation, and self-efficacy (Carolus et al., 2023; Long & Magerko, 2020; Markus et al., 2025; Ng et al., 2021; Nong et al., 2024).

The results show that AI Ethics is the only indicator that differs significantly between the two class groups. This suggests that students who are more frequently exposed to technology-oriented learning environments may develop stronger awareness of ethical issues related to AI, such as privacy, algorithmic bias, and responsible use (Baker & Hawn, 2022; Muraina & Adesanya, 2024; Walter, 2024; Zhang et al., 2024). From a practical perspective, this finding highlights the importance of integrating ethical discussions about AI within school learning activities. Schools can strengthen AI literacy not only by introducing AI tools but also by embedding ethical reflection in digital literacy or technology-related subjects. For example, classroom activities may involve discussing real cases of algorithmic bias, data privacy risks, or responsible AI use in everyday digital applications. Such approaches may help students develop a more critical and responsible perspective toward AI technologies. However, because this study used a comparative ex post facto design with intact classes, the findings should be interpreted as non-causal differences between groups rather than direct effects of instructional treatment (Expósito López, 2020; Sharma & Adhikari, 2022). The results therefore suggest possible contextual influences related to technology-based learning environments, which require further investigation in future studies.

In contrast, the Apply/Use indicator did not show a statistically significant difference between IT and regular classes. The absence of differences in the Apply/Use indicator suggests that students' experience with AI tools may extend beyond classroom contexts. AI usage skills may therefore develop through broader exposure to digital devices, online platforms, and everyday technology use outside the classroom. This finding supports the view that different dimensions of AI literacy may develop unevenly depending on learning contexts (Carolus et al., 2023; Markus et al., 2025; Nong et al., 2024).

Similarly, the Attitude/Motivation/Self-efficacy and AI_Total indicators did not show statistically significant differences between the two classes. The relatively similar scores suggest that students in both classes have comparable levels of confidence and motivation when engaging with AI technologies. These findings suggest that strengthening AI literacy in schools should not be limited to specific academic tracks such as IT classes. Instead, school-wide programs that integrate AI literacy within broader digital literacy initiatives may help ensure that all students develop balanced competencies, particularly in areas such as ethical awareness and responsible AI use (Walter, 2024; Zhang et al., 2024).

For the Understand indicator, which was measured using a dichotomous test (0–5), no statistically significant difference was found between classes. However, this result should be interpreted cautiously because the indicator showed relatively low reliability and limited response variation, which may reduce the sensitivity of the instrument’s sensitivity in detecting class differences. Therefore, the absence of differences in the Understand indicator should not be interpreted as definitive evidence that the conceptual understanding of the two classes is identical, but rather as a limitation of the measurement instrument used in this study. In line with the focus of the adapted AI literacy instrument (P. Kesuma & Amelia Fransen, 2025), further research could test the hypothesis that conceptual understanding measured with more varied items (greater number and more balanced difficulty levels) will show a clearer correlation with the ethics and usage indicators.

When linked to the research background that has been presented, these findings are also relevant to filling the gaps in empirical evidence on AI literacy at the junior high school level, while also showing that differences between IT and regular classes do not automatically appear in all aspects of AI literacy (Chiu et al., 2024; Israwati Hamsar et al., 2024; P. Kesuma & Amelia Fransen, 2025). Several limitations should also be considered when interpreting these findings. First, the ex post facto design with intact classes limits causal interpretation and allows the possibility of confounding (Expósito López, 2020; Sharma & Adhikari, 2022). Second, the sample coverage in one school limits generalization. Third, self-report indicators may be affected by response bias and limited score variation. Fourth, the Understand indicator has low reliability due to limited items and unbalanced difficulty levels. Consequently, the finding of “no difference” in Understand should be understood as not yet adequately detected by the current instrument (low sensitivity), rather than as strong evidence that the conceptual abilities of the two groups are completely the same. Future research could address these limitations by employing longitudinal or quasi-experimental designs with baseline measurements, expanding the sample across multiple schools, and refining the measurement instruments, particularly by improving the number and difficulty balance of items in the Understand indicator. Further studies may also explore whether exposure to technology-based learning environments contributes to the development of AI ethics awareness and whether targeted AI ethics instruction can promote more balanced AI literacy development across different class groups (Walter, 2024; Zhang et al., 2024).

4. Conclusion

This study examined that junior high school students' AI literacy as a multidimensional construct that includes conceptual understanding (Understand), application skills (Apply/Use), ethical considerations (AI Ethics), and attitude/motivation/self-efficacy. Overall, the Likert scale–based indicators show relatively positive profiles, indicating that students generally demonstrate readiness to engage with AI technologies in learning contexts. The comparison between IT classes and regular classes shows that statistically significant differences appear only in the AI Ethics dimension, while the Apply/Use, Attitude/Motivation/Self-efficacy, and total Likert scores remain relatively similar between the two groups. These findings highlight the importance of strengthening AI literacy programs that address multiple dimensions, particularly ethical awareness, so that AI-related learning opportunities can be developed more equitably across different class groups. Meanwhile, the results for the Understand indicator should be interpreted cautiously due to measurement limitations, particularly the relatively low reliability and limited variation of test items. Therefore, the findings cannot yet serve as strong evidence regarding differences in conceptual understanding between the two classes. Overall, this study provides empirical insights into the profile of AI literacy among junior high school students and contributes practical implications for schools in integrating AI literacy into learning activities. Future research may expand this work by refining the measurement of conceptual understanding and examining how exposure to technology-based learning environments influences the development of AI ethics awareness.

References

Antoniou, F., & Alghamdi, M. H. (2024). Confidence in mathematics is confounded by responses to reverse-coded items. *Frontiers in Psychology*, *15*(October). https://doi.org/10.3389/fpsyg.2024.1489054

Baker, R. S., & Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, *32*(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, *1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

Chicco, D., Sichenze, A., & Jurman, G. (2025). A simple guide to the use of Student's t-test, Mann-Whitney U test, Chi-squared test, and Kruskal-Wallis test in biostatistics. *BioData Mining*, *18*(1). https://doi.org/10.1186/s13040-025-00465-6

Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, *6*(March), 100171. https://doi.org/10.1016/j.caeo.2024.100171

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*(1), 1–17. https://doi.org/10.1186/2193-1801-2-222

Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1. https://doi.org/10.11648/j.ajtas.20160501.11

Fauziah Wulansari, Riska Hartanti, Bambang Sumardjoko, S. N. (2026). The Implementation of AI-Assisted Digital Artwork Learning as an Innovative Strategy for Elementary School Teachers in Improving Student's Technological Literacy in the Era of the Independent Curriculum. *Journal of Educational Sciences*, *10*. https://doi.org/https://doi.org/10.31258/jes.10.1.p.2113-2127

Israwati Hamsar, Hajar Dewantara, Hijri Andini, Sulfan Ardiansyah, Nur Afrilia Afrizah, & Muh Getri Putra Hasta. (2024). Analisis Literasi Artifical Intelligence Mahasiswa Pada Perguruan Tinggi. *Journal of Vocational, Informatics and Computer Education*, *2*(1), 72–81.

https://doi.org/10.61220/voice.v2i1.31

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376727

López, J. E., Chacón-Cuberos, R., Parra-González, M. E., Aguaded-Ramírez, E. M., & Lacárcel, A. C. (2020). Tutorial action and emotional development of students as elements of improved development and preventing problems related with coexistence and social aspects. *European Journal of Investigation in Health, Psychology and Education*, *10*(2), 615–627. https://doi.org/10.3390/ejihpe10020045

Markus, A., Carolus, A., & Wienrich, C. (2025). Objective measurement of AI literacy: Development and validation of the AI competency objective scale (AICOS). *Computers and Education: Artificial Intelligence*, *9*. https://doi.org/10.1016/j.caeai.2025.100485

Muraina, I. O., & Adesanya, M. O. (2024). Impact of AI-ChatGPT Intervention on Coding: NPL Supportive Approach to Teaching and Learning Effectiveness. *Journal of Educational Sciences*, *8*(3), 312–324. https://doi.org/10.31258/jes.8.3.p.312-324

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nong, Y., Cui, J., He, Y., Zhang, P., & Zhang, T. (2024). Development and Validation of an AI Literacy Scale. *Journal of Artificial Intelligence Research*, *1*(1), 17–26. https://doi.org/10.70891/jair.2024.100029

P. Kesuma, D., & Amelia Fransen, L. (2025). Implementasi Instrumen MAILS (Meta AI Literacy Scale) Untuk Pengukuran Tingkat Literasi AI Pada Mahasiswa Ilmu Komputer. *Jurnal Sistem Informasi, Manajemen Dan Teknologi Informasi*, *3*(1), 25–36. https://doi.org/10.33020/jsimtek.v3i1.811

Rochon, J., Gondan, M., & Kieser, M. (2012). To test or not to test: Preliminary assessment of normality when comparing two independent samples. *BMC Medical Research Methodology*, *12*. https://doi.org/10.1186/1471-2288-12-81

Rosiana, Y., Syafii, W., Putra, R. A., & Futra, D. (2023). Effectiveness of E-Module Circulation System Based on 8E Learning Cycle on High School Student's Critical Thinking Ability. *Journal of Educational Sciences*, *7*(3), 434. https://doi.org/10.31258/jes.7.3.p.434-441

Samiha, Y. T., & Aksara, L. (2025). PGMI Students' Perceptions of the Implementation of Artificial Intelligence (AI) Technology in Learning in Madrasah Ibtidaiyah. *Journal of Educational Sciences*, *9*(3), 1055–1065.

Sharma, M. K., & Adhikari, R. (2022). Effects of Water, Sanitation, and Hygiene on the School Absenteeism of Basic Level Students in the Government School of Nepal. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.869933

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. In *International journal of medical education* (Vol. 2, pp. 53–55). https://doi.org/10.5116/ijme.4dfb.8dfd

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom:

the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00448-3

Wardhani, I. S., & Asyiah, I. N. (2025). *AI (Artificial Intelligence) -Enhanced Digital Literacy in Science*. *9*(4), 1957–1970. https://doi.org/10.31258/jes.9.4.p.1957-1970

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? In *International Journal of Educational Technology in Higher Education* (Vol. 16, Issue 1). https://doi.org/10.1186/s41239-019-0171-0

Zhang, H., Lee, I., & Moore, K. (2024). An Effectiveness Study of Teacher-Led AI Literacy Curriculum in K-12 Classrooms. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), 23318–23325. https://doi.org/10.1609/aaai.v38i21.30380

How to cite this article:

Rahayu, S., & Elviani, U. (2026). Analysis of Differences in AI Literacy Levels among Students at SMPN 1 Purwakarta in IT Classes and Regular Classes. *Journal of Educational Sciences*, *10(4)*, 910-926.