

## THE EFFECT OF SELF-EFFICACY ON DATA LITERACY SKILLS AMONG JUNIOR HIGH SCHOOL STUDENTS IN THE ULU SOSA SUBDISTRICT

Yulia Nasution<sup>1</sup>, Suci Dahlya Narpila<sup>2</sup>

<sup>1,2</sup> Universitas Islam Negeri Sumatera Utara, Indonesia

e-mail: [yulianst06@gmail.com](mailto:yulianst06@gmail.com)

### Abstract

Data literacy is an essential competency that enables students to read, interpret, analyze, evaluate, and draw conclusions from information presented in various forms of data. Self-efficacy may be associated with students' engagement and persistence when dealing with data-based tasks. This study aimed to examine the statistical contribution of self-efficacy to the data literacy skills of eighth-grade students in Ulu Sosa Subdistrict. The study employed a quantitative approach with an ex post facto design. The participants consisted of 30 eighth-grade students from one junior high school in Ulu Sosa Subdistrict. Data were collected using a self-efficacy questionnaire and a five-item data literacy essay test. The data were analyzed using descriptive statistics, assumption testing, and simple linear regression. The results showed a positive and statistically significant association between self-efficacy and students' data literacy skills, with a significance value of 0.001. The regression model produced an  $R^2$  value of 0.948, indicating that 94.8% of the variance in data literacy scores within the sample was statistically accounted for by self-efficacy. However, because the study employed an ex post facto design and involved only 30 students from one school, the finding should not be interpreted as evidence of a causal relationship or generalized to all junior high school students in Ulu Sosa Subdistrict. The findings indicate a strong statistical association between self-efficacy and data literacy skills and provide a basis for further experimental or longitudinal research.

**Keywords:** self-efficacy, data literacy, junior high school students

### Abstrak

Literasi data merupakan kompetensi penting yang memungkinkan siswa membaca, menginterpretasikan, menganalisis, mengevaluasi, dan menarik kesimpulan berdasarkan informasi yang disajikan dalam berbagai bentuk data. Self-efficacy diduga berkaitan dengan keterlibatan dan ketekunan siswa ketika menghadapi tugas berbasis data. Penelitian ini bertujuan untuk menganalisis kontribusi secara statistik self-efficacy terhadap kemampuan literasi data siswa kelas VIII di Kecamatan Ulu Sosa. Penelitian ini menggunakan pendekatan kuantitatif dengan desain ex post facto. Partisipan penelitian terdiri atas 30 siswa kelas VIII dari satu sekolah menengah pertama di Kecamatan Ulu Sosa. Data dikumpulkan menggunakan angket self-efficacy dan tes uraian kemampuan literasi data yang terdiri atas lima butir soal. Data dianalisis menggunakan statistik deskriptif, uji prasyarat, dan regresi linear sederhana. Hasil penelitian menunjukkan adanya hubungan positif dan signifikan secara statistik antara self-efficacy dan kemampuan literasi data siswa dengan nilai signifikansi sebesar 0,001. Model regresi menghasilkan nilai  $R^2$  sebesar 0,948, yang menunjukkan bahwa 94,8% variasi skor kemampuan literasi data dalam sampel secara statistik dapat dijelaskan oleh self-efficacy. Namun, karena penelitian menggunakan desain ex post facto dan hanya melibatkan 30 siswa dari satu sekolah, hasil tersebut tidak dapat ditafsirkan sebagai bukti hubungan sebab-akibat dan tidak dapat digeneralisasikan kepada seluruh siswa SMP di Kecamatan Ulu Sosa. Temuan penelitian menunjukkan adanya hubungan statistik yang kuat antara self-efficacy dan kemampuan literasi data serta memberikan dasar bagi penelitian eksperimental atau longitudinal selanjutnya.

**Kata kunci:** *self-efficacy*, kemampuan literasi data, siswa SMP

### INTRODUCTION

The development of information technology in the digital age has transformed the way individuals acquire, manage, and utilize information in their daily lives. Various decisions in

the fields of education, economics, health, and social affairs are increasingly based on data presented in the form of tables, graphs, diagrams, and digital visualizations. This situation requires every individual to possess data literacy as one of the essential competencies of the 21st century. Data literacy is not only related to the ability to read data but also encompasses the ability to understand, interpret, analyze, evaluate, and draw conclusions based on available information (Gal, 2002; Schield, 2004). Therefore, the development of data literacy skills needs to be a priority in the learning process at school, particularly in mathematics classes.

The urgency of improving data literacy skills is becoming increasingly evident through various international assessment results. The International Computer and Information Literacy Study (ICILS) report indicates that students' ability to understand and use data-based information still requires serious attention in order to navigate the digital society (Fraillon et al., 2023). Furthermore, a report by the Organization for Economic Co-operation and Development (OECD) emphasizes that students need the ability to interpret information, think critically, and use data as a basis for decision-making in order to adapt to advancements in science and technology (OECD, 2021). In line with these policies, the Indonesian government, through the Merdeka Curriculum, has also identified critical thinking, problem-solving, and literacy as competencies that must be developed in the learning process (Ministry of Education and Culture, 2020).

Nevertheless, various studies indicate that students' data literacy and mathematical literacy skills are still not optimal. Students are generally able to read information presented in the form of tables or graphs, but they still struggle to interpret relationships between data, evaluate information, and draw appropriate conclusions (Haqqul et al., 2022). This situation indicates that mastery of mathematical concepts alone is not sufficient to develop strong data literacy skills. Internal factors are needed to encourage students to be confident, persevere in the face of difficulties, and strive to solve problems related to data analysis.

One psychological factor believed to influence data literacy is self-efficacy. Bandura (1997) defines self-efficacy as an individual's belief in their ability to organize and carry out the actions necessary to achieve a specific goal. Individuals with high self-efficacy tend to have greater motivation to learn, are willing to face challenges, do not give up easily, and are able to use learning strategies effectively. Schunk and DiBenedetto (2020) also explain that self-

efficacy plays a role in enhancing motivation, perseverance, and higher-order thinking skills, thereby impacting students' learning success.

The relationship between self-efficacy and literacy skills has been extensively studied in various research studies. Indrawati and Wardono (2019) found that self-efficacy has a positive effect on students' mathematical literacy skills. A study by Khotimah et al. (2020) also showed that students with high levels of self-efficacy achieved better mathematics learning outcomes than students with low levels of self-efficacy. Furthermore, Kong et al. (2022) explained that self-efficacy is a key factor in the development of data literacy because it boosts students' confidence in understanding, processing, and utilizing data-based information.

Previous studies have shown that self-efficacy is associated with students' mathematical achievement and mathematical literacy. However, an important issue that remains insufficiently examined concerns the relationship between self-efficacy and the specific processes involved in data literacy. Data literacy requires students not only to read numerical information but also to interpret relationships among data, analyze patterns, evaluate information, and formulate conclusions based on evidence.

These processes require students to remain engaged and persistent when dealing with complex information. However, the extent to which self-efficacy is statistically associated with students' ability to perform these data literacy processes remains insufficiently examined, particularly among junior high school students. Therefore, this study addresses this empirical gap by analyzing the statistical contribution of self-efficacy to students' data literacy skills without assuming that self-efficacy directly causes improvements in data literacy.

Based on preliminary observations at the Ulu Sosa Subdistrict Integrated Public Junior High School, it was found that some students still struggle to read, interpret, and analyze data presented in tables or graphs. Additionally, some students expressed doubts about their abilities when solving problems that require data-based reasoning. These conditions indicate that the psychological aspect of self-efficacy is likely to influence students' data literacy skills.

In an effort to provide empirical evidence regarding this issue, this study analyzes the influence of self-efficacy on students' data literacy skills using a quantitative approach through simple linear regression analysis. The results of this study are expected to provide information on the extent to which self-efficacy contributes to data literacy skills and to serve

as a basis for teachers in designing learning strategies that not only enhance cognitive abilities but also strengthen students' self-confidence in solving data-based problems.

Based on the foregoing discussion, this study aimed to analyze the statistical contribution of self-efficacy to the data literacy skills of eighth-grade students in the research sample in Ulu Sosa Subdistrict.

## METHODS

This study employed a quantitative approach with an ex post facto design. This design was selected because the researcher did not provide an intervention or manipulate the self-efficacy variable. Self-efficacy and data literacy skills were measured as they naturally occurred among the participants. Simple linear regression was subsequently used to analyze the statistical relationship and predictive contribution of self-efficacy to data literacy scores. Therefore, the regression results are interpreted as statistical association and contribution rather than evidence of a causal relationship.

The study was conducted during the second semester of the 2025/2026 academic year at a junior high school in Ulu Sosa Subdistrict, Padang Lawas Regency, North Sumatra. The research site was selected based on preliminary observations indicating variation in students' data literacy skills and the limited empirical evidence specifically examining the statistical relationship between self-efficacy and data literacy within this context.

The population consisted of 113 eighth-grade students distributed across three junior high schools in Ulu Sosa Subdistrict: SMP Negeri 1 Ulu Sosa, SMP Negeri Satu Atap Ulu Sosa, and SMP Swasta Nurul Iman Ulu Sosa. The sample consisted of 30 students selected using cluster random sampling. The 30 participants in the study were students from SMP Negeri Satu Atap Ulu Sosa. Therefore, the findings describe the relationship observed within this sample and cannot be directly generalized to all junior high school students in Ulu Sosa Subdistrict without further research involving a larger and more representative sample.

The research subjects were the eighth-grade students who served as the study sample. The research variables consisted of two variables: self-efficacy as the independent variable (X) and data literacy skills as the dependent variable (Y). The data used in this study consisted of primary data-data obtained directly from respondents through the distribution of self-efficacy questionnaires and the administration of data literacy tests. In addition, secondary

data in the form of information regarding the number of students, school profiles, and other supporting documents obtained from the school were used.

The research instruments consisted of a self-efficacy questionnaire and a data literacy test. The self-efficacy questionnaire was designed based on the three dimensions proposed by Bandura: magnitude, strength, and generality. The questionnaire comprised 12 items, including 6 positive statements and 6 negative statements. Measurements used a four-point Likert scale: Strongly Agree (SS), Agree (S), Disagree (TS), and Strongly Disagree (STS). Scores for positive statements were assigned sequentially as 4, 3, 2, and 1, while scores for negative statements were assigned in reverse order.

Data literacy was measured using an essay test consisting of five items. The items were developed based on five data literacy indicators: (1) reading data, (2) interpreting data, (3) analyzing relationships or patterns among data, (4) evaluating information presented in data, and (5) drawing conclusions based on evidence. Each item was designed to represent one primary indicator. The test was scored using a rubric that considered the accuracy, completeness, and relevance of students' responses to the data presented. Before being used in the study, the instrument underwent expert content validation and empirical validity testing. The reliability of the data literacy test was analyzed using Cronbach's Alpha and produced a coefficient of 0.760713.

The data collection process was conducted in several stages. The first stage involved initial observations to gain an understanding of the learning environment and students' data literacy skills. The second stage involved distributing self-efficacy questionnaires to all respondents in the sample classes. After the questionnaires were collected, the next stage involved administering a data literacy test to the respondents within a specified timeframe. All data collection activities were conducted in person at the school while adhering to the learning schedule agreed upon with school officials.

The collected data were analyzed using IBM SPSS Statistics version 25. The analysis began with descriptive statistics to describe the data characteristics through the mean, standard deviation, minimum value, and maximum value. Next, prerequisite tests for analysis were conducted, including a normality test using the Shapiro–Wilk test and a linearity test using the Test for Linearity. Hypothesis testing was performed using simple linear regression analysis to determine the effect of self-efficacy on students' data literacy skills.

The regression model used was:

$$Y = a + bX$$

Where Y represents data literacy, X represents self-efficacy, a is a constant, and b is the regression coefficient. Hypothesis testing decisions are based on a 5% significance level ( $\alpha = 0.05$ ). The research hypothesis is accepted if the significance value is less than 0.05. In addition, the magnitude of the contribution of the self-efficacy variable to data literacy ability is analyzed using the coefficient of determination ( $R^2$ ) to determine the percentage of variation in data literacy ability that can be explained by the self-efficacy variable.

## RESULTS AND DISCUSSION

### Results

This study was conducted to determine the effect of self-efficacy on the data literacy skills of junior high school (SMP) students in Ulu Sosa Subdistrict. The study employed a quantitative approach using simple linear regression analysis to determine the effect and magnitude of the influence of self-efficacy as the independent variable on data literacy skills as the dependent variable.

Research data were collected from 30 eighth-grade students at the Ulu Sosa Subdistrict Integrated Public Junior High School. Data collection was conducted using two research instruments: a self-efficacy questionnaire and a data literacy test. The self-efficacy questionnaire was designed based on indicators of task difficulty (magnitude), strength of belief (strength), and scope of the task (generality). Meanwhile, the data literacy test was designed based on indicators of the ability to read, understand, interpret, analyze, and draw conclusions based on data. Before the instruments were used in the study, their quality was assessed through expert validation, validity testing, and reliability testing. These steps were taken to ensure that the instruments used could measure the research variables accurately and consistently.

### Results of the Research Instrument Validation

The instrument was validated by two validators: a mathematics education professor and a middle school mathematics teacher. The validation assessed the content appropriateness, instrument construction, and clarity of language used in both the questionnaire and the data literacy test.

The results of the expert validation of the research instrument are presented in Table 1.

Table 1. Results of the Research Instrument Validation

| No | Instrument                   | Validator                                | Total Score | Number of Statements | Average Score | Criteria     |
|----|------------------------------|------------------------------------------|-------------|----------------------|---------------|--------------|
| 1. | Self- Efficacy Questionnaire | Mathematics Education Faculty            | 44          | 12                   | 3,67          | Highly Valid |
|    |                              | Junior High School Mathematics Education | 42          | 12                   | 3,50          | Highly Valid |
| 2. | Data Literacy Test           | Mathematics Education Faculty            | 63          | 17                   | 3,65          | Highly Valid |
|    |                              | Junior High School Mathematics Education | 63          | 17                   | 3,71          | Highly Valid |

Based on Table 1, it can be seen that all instruments were classified as highly valid. These results indicate that the instruments met the validity criteria based on expert assessment, in terms of indicator appropriateness, item presentation, and language use. Thus, the self-efficacy questionnaire and the data literacy test can be used in the research data collection phase.

### Instrument Validity Test Results

Validity testing was conducted to determine the instrument's accuracy in measuring the research variables. Validity testing was performed using the product-moment correlation by comparing the calculated  $r$  value with the table  $r$  value. The instrument was deemed valid if the calculated  $r$  value  $>$  table  $r$  value. In this study, the table  $r$  value used was 0.349.

The critical  $r$  value was determined based on the number of respondents used in the validity test, the degrees of freedom ( $n - 2$ ), and the selected significance level. Therefore, the critical  $r$  value should be verified based on the actual number of respondents involved in the validity testing. The study should also clarify whether the empirical validity test was conducted using an independent pilot sample or the same respondents as those included in the main study.

### Data Literacy Test Results

---

*The Effect of Self-Efficacy on Data Literacy Skills Among Junior High School Students in the Ulu Sosa Nasution, Narpila*

The results of the validity test for the data literacy test, which consists of five items, show that all items meet the validity criteria.

**Table 2. Results of the Data Literacy Test Validity Test**

| Variable              | Item | Calculated r | Table r | Notes |
|-----------------------|------|--------------|---------|-------|
| Ability Data Literacy | P1   | 0,485358     | 0,349   | Valid |
|                       | P2   | 0,361312     | 0,349   | Valid |
|                       | P3   | 0,448458     | 0,349   | Valid |
|                       | P4   | 0,534825     | 0,349   | Valid |
|                       | P5   | 0,571053     | 0,349   | Valid |

Based on Table 2, all items obtained a calculated r value greater than the table r value. This indicates that each item is capable of accurately measuring students' data literacy skills, so all items can be used in the study.

#### **Validity of the Self-Efficacy Questionnaire**

A validity test was also conducted on the self-efficacy questionnaire, which consists of 12 statements.

**Table 3. Results of the Self-Efficacy Questionnaire Validity Test**

| Variable      | Item | Calculated r | Table r | Notes |
|---------------|------|--------------|---------|-------|
| Self-efficacy | P1   | 0,413121     | 0,349   | Valid |
|               | P2   | 0,435935     | 0,349   | Valid |
|               | P3   | 0,415828     | 0,349   | Valid |
|               | P4   | 0,508269     | 0,349   | Valid |
|               | P5   | 0,578197     | 0,349   | Valid |
|               | P6   | 0,59364      | 0,349   | Valid |
|               | P7   | 0,608515     | 0,349   | Valid |
|               | P8   | 0,551369     | 0,349   | Valid |
|               | P9   | 0,640483     | 0,349   | Valid |
|               | P10  | 0,585874     | 0,349   | Valid |
|               | P11  | 0,583948     | 0,349   | Valid |
|               | P12  | 0,381722     | 0,349   | Valid |

The test results show that all statements in the self-efficacy questionnaire have a calculated r value > table r value. Thus, all items are deemed valid and can be used to measure students' levels of self-efficacy.

### Results of the Instrument Reliability Test

After the instrument was deemed valid, the next step was to conduct a reliability test to determine the instrument's consistency in measuring the research variables. A reliable instrument indicates that it is capable of producing relatively consistent measurement results when used under the same conditions.

In this study, reliability tests were conducted on the data literacy test and the self-efficacy questionnaire. The results of the instrument reliability tests are presented in Table 4.

**Table 4. Results of the Research Instrument Reliability Tests**

| Instrument                    | Reliability Test Results | Interpretation | Criteria |
|-------------------------------|--------------------------|----------------|----------|
| Data Literacy Test            | 0,760713                 | Reliable       | High     |
| Self – Efficacy Questionnaire | 0,752673                 | Reliable       | High     |

Based on Table 4, the reliability coefficient for the data literacy test was 0.760713, and the reliability coefficient for the self-efficacy questionnaire was 0.752673. Both values fall into the high category; therefore, it can be concluded that the research instruments demonstrate a good level of consistency and are suitable for use as research data collection tools.

### Results of Descriptive Statistical Analysis

Descriptive statistical analysis was conducted to provide an overview of the characteristics of the research data. The results of the analysis are presented in Table 5.

**Table 5. Descriptive Statistics of Research Variables**

| Variable             | Number of Student | Maximum | Minimum | Mean  | Median | Standar deviation |
|----------------------|-------------------|---------|---------|-------|--------|-------------------|
| Data Literacy Skills | 30                | 95      | 64      | 80,27 | 80,00  | 8,816             |
| Self - efficacy      | 30                | 48      | 27      | 38,27 | 38,00  | 5,771             |

Based on the results of the descriptive statistical analysis, an overview of the characteristics of the data regarding the variables of students' self-efficacy and data literacy skills was obtained. This analysis aimed to identify trends in the data prior to conducting inferential analysis. The results showed that students' self-efficacy scores had a mean of

38.27, with a minimum score of 27 and a maximum score of 48. The standard deviation of 5.771 describes the dispersion of individual scores around the mean. A standard deviation smaller than the mean cannot be used as evidence that the data are representative or that the distribution is sufficiently narrow. Therefore, the descriptive statistics are used to describe the characteristics of self-efficacy scores obtained from the research sample.

Meanwhile, the data literacy scores had a mean of 80.27, with a minimum score of 64 and a maximum score of 95. The standard deviation of 8.816 indicates the dispersion of data literacy scores around the mean. The interpretation of the data literacy category should be based on predetermined category intervals and should be explained in the study. Therefore, the mean score is used to describe the tendency of data literacy scores within the sample rather than as the sole basis for assigning a particular category to all students. Thus, the data obtained are sufficiently representative for use in further analysis.

### Prerequisite Tests for Analysis

Simple linear regression analysis was conducted to examine the statistical relationship and predictive contribution of self-efficacy to students' data literacy skills.

### Normality Test

The normality test aims to determine whether the research data come from a normal distribution. In this study, the normality test was conducted using the Shapiro-Wilk test with the aid of SPSS software at a significance level of 0.05. The results of the normality test are presented in Table 6.

**Table 6. Results of the Normality Test**

|                             | Kolmogorov-Smirnov <sup>a</sup> |    |                   | Shapiro-Wilk |    |      |
|-----------------------------|---------------------------------|----|-------------------|--------------|----|------|
|                             | Statistic                       | df | Sig.              | Statistic    | df | Sig. |
| ANGKET SELF EFFICACY        | .115                            | 30 | .200 <sup>*</sup> | .966         | 30 | .448 |
| TES KEMAMPUAN LITERASI DATA | .092                            | 30 | .200 <sup>*</sup> | .968         | 30 | .498 |

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

a. Lilliefors Significance Correction

The p-values for each variable were greater than 0.05, so it can be concluded that both the self-efficacy and data literacy ability data follow a normal distribution. Thus, one of the assumptions of regression analysis has been met.

### Linearity Test

A linearity test was conducted to determine whether the relationship between self-efficacy (the independent variable) and data literacy skills (the dependent variable) is linear. The results of the linearity test for the relationship between self-efficacy and data literacy skills are presented in Table 7.

**Table 7. Linearity Test Results**

| ANOVA Table                                        |                |                          | Sum of Squares | df | Mean Square | F       | Sig.   |
|----------------------------------------------------|----------------|--------------------------|----------------|----|-------------|---------|--------|
| TES KEMAMPUAN LITERASI DATA * ANGKET SELF EFFICACY | Between Groups | (Combined)               | 2195.700       | 15 | 146.380     | 35.232  | <,.001 |
|                                                    |                | Linearity                | 2137.547       | 1  | 2137.547    | 514.481 | <,.001 |
|                                                    |                | Deviation from Linearity | 58.153         | 14 | 4.154       | 1.000   | .500   |
|                                                    | Within Groups  |                          | 58.167         | 14 | 4.155       |         |        |
|                                                    | Total          |                          | 2253.867       | 29 |             |         |        |

The significance value for Deviation from Linearity is 0.328, which is greater than 0.05; therefore, the relationship between self-efficacy and data literacy skills is considered linear. Consequently, the data meet the criteria for analysis using simple linear regression.

### Simple Linear Regression Analysis

A simple linear regression analysis was conducted to determine the extent of the effect of self-efficacy on students' data literacy skills. The results of the analysis are shown in Table 8.

**Table 8. Results of the Simple Linear Regression Analysis**

| Coefficients <sup>a</sup> |                             |            |                           |       |        |
|---------------------------|-----------------------------|------------|---------------------------|-------|--------|
| Model                     | Unstandardized Coefficients |            | Standardized Coefficients | t     | Sig.   |
|                           | B                           | Std. Error | Beta                      |       |        |
| 1                         | (Constant)                  | 23.339     | 2.537                     | 9.199 | <,.001 |
|                           | ANGKET SELF EFFICACY        | 1.488      | .066                      | .974  | <,.001 |

a. Dependent Variable: TES KEMAMPUAN LITERASI DATA

Based on the results of the analysis, the following regression equation was obtained:

$$Y = 23.339 + 1.488 X$$

This equation indicates that the constant of 23.339 represents the data literacy ability score when the self-efficacy score is held constant. The regression equation produced a regression coefficient of 1.488. The positive coefficient indicates a positive statistical association between self-efficacy and data literacy scores within the research sample. In other words, higher self-efficacy scores tended to be accompanied by higher data literacy scores.

This finding represents a statistical association rather than evidence that an increase in self-efficacy directly causes an increase in data literacy skills.

The significance value of 0.001 was lower than the significance level of 0.05. Thus, a positive and statistically significant association was found between self-efficacy and data literacy skills within the research sample.

### Coefficient of Determination

The extent of self-efficacy's contribution to data literacy skills was analyzed using the coefficient of determination. The results of the analysis are presented in Table 9.

**Table 9. Coefficient of Determination Results**

| Model Summary |                   |          |                   |                            |
|---------------|-------------------|----------|-------------------|----------------------------|
| Model         | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1             | .974 <sup>a</sup> | .948     | .947              | 2.038                      |

a. Predictors: (Constant), ANGKET SELF EFFICACY

The R Square value of 0.948 indicates that 94.8% of the variance in data literacy scores within the sample was statistically accounted for by self-efficacy in the regression model. This value is exceptionally high and therefore requires cautious interpretation. The  $R^2$  value should not be interpreted as evidence that self-efficacy causally determines 94.8% of data literacy skills or that other factors causally account for only 5.2%. The magnitude of  $R^2$  should be considered a characteristic of the statistical model fitted to the present sample.

### Interpretation of Research Results

Based on the results of the statistical analysis, it is evident that self-efficacy has a positive and significant effect on students' data literacy skills. The higher students' level of confidence in their abilities, the greater their ability to understand, interpret, analyze, evaluate, and draw conclusions based on data.

The high value of the coefficient of determination (94.8%) indicates that self-efficacy is a highly dominant factor in explaining the variation in data literacy skills among the study sample. These results suggest that students who have high confidence in their abilities tend to be more self-assured in solving problems that require data-based reasoning, more persistent when facing difficulties, and better able to select appropriate problem-solving strategies.

Overall, the analysis results show that self-efficacy has a positive influence on students' data literacy skills. Students with higher levels of self-efficacy tend to achieve better scores

on data literacy skills. These findings are evidenced by positive regression coefficients and a high coefficient of determination. Thus, the research results provide empirical evidence that enhancing self-efficacy has the potential to improve students' ability to understand, interpret, analyze, evaluate, and draw conclusions based on data presented in various forms of representation.

## Discussion

Based on the simple linear regression analysis, a regression coefficient of 1.488 was obtained with a significance value of 0.001. The result indicates a positive and statistically significant association between self-efficacy and students' data literacy skills within the research sample. The positive regression coefficient indicates that students with higher self-efficacy scores tended to have higher data literacy scores. However, because the study employed an ex post facto design, this association cannot be interpreted as a causal relationship.

This association can be understood through the characteristics of the processes involved in data literacy. Reading data requires attention to numerical and visual information, whereas interpretation requires students to connect information with the context of a problem. Analysis and evaluation require persistence, comparison of evidence, and consideration of possible interpretations. Drawing conclusions subsequently requires the ability to make judgments based on available evidence. Higher self-efficacy may theoretically be associated with students' willingness to engage in these processes, persist when difficulties arise, and review their solutions. Thus, the relationship between self-efficacy and data literacy may be associated with persistence, strategic engagement, and self-regulation during data-based tasks.

On the other hand, the coefficient of determination of 94.8% obtained in this study is exceptionally high and exceeds values reported in several previous studies. This value requires critical interpretation because the study involved only 30 students and used a single predictor. In addition, the possibility that particular observations exerted substantial influence on the regression model should be examined empirically. Therefore, the relatively small or homogeneous sample cannot be directly considered an explanation for the high  $R^2$  without further examination. The value is more appropriately viewed as a characteristic of

the statistical model fitted to the present sample. Future research using larger and more diverse samples is needed to determine whether such a strong relationship can be replicated.

The findings of this study can be explained through Social Cognitive Theory, as proposed by Bandura (1997). According to Bandura, self-efficacy is an individual's belief in their ability to organize and carry out the actions necessary to achieve a specific outcome. This belief influences how individuals think, behave, motivate themselves, and persevere when facing difficulties. In the context of mathematics learning, students with high self-efficacy tend to be more confident when facing problems that require reasoning, are more willing to try various alternative solutions, and do not give up easily when encountering difficulties. Conversely, students with low self-efficacy tend to avoid tasks they consider difficult, quickly feel discouraged, and lack confidence in their abilities, which results in low data literacy skills.

Data literacy is a complex skill because it involves several sequential thought processes, ranging from reading data, understanding information, interpreting relationships between data points, conducting analysis, evaluating information, to drawing conclusions based on available evidence. According to Gal (2002), data literacy requires not only the ability to read numbers but also the ability to use information critically in decision-making. Schield (2004) also explains that data literacy includes the ability to understand the context of data and evaluate the quality of information before using it as a basis for decision-making. Therefore, students need confidence in their own abilities to be able to complete each of these stages optimally.

The findings of this study also indicate that students with high self-efficacy tend to be more active in solving problems that require data analysis. They are not only able to interpret information presented in tables or graphs but are also capable of explaining relationships between variables, identifying patterns in the data, and drawing conclusions based on the available evidence. Conversely, students with low self-efficacy experience more difficulty when they have to connect information from various data sources, resulting in a less optimal problem-solving process. These findings indicate that the psychological aspect of self-efficacy plays a crucial role in supporting the higher-order thinking processes required for data literacy.

The results of this study align with the research by Indrawati and Wardono (2019), which states that self-efficacy has a positive effect on students' mathematical literacy skills.

That study showed that students with high confidence in their abilities were better able to understand problems, determine solution strategies, and evaluate their work independently. These similar findings indicate that self-efficacy is an internal factor that contributes to students' success in developing various forms of literacy, including data literacy.

These findings are also consistent with the study by Khotimah et al. (2020), which found that self-efficacy has a positive relationship with mathematics learning outcomes. According to that study, students who believe in their own abilities tend to have higher motivation to learn, are more persistent in completing tasks, and are able to maintain their efforts when facing difficulties. These characteristics are essential for solving data literacy problems, which generally require critical and analytical thinking skills.

Furthermore, the results of this study support the research by Kong et al. (2022), which explains that self-efficacy is one of the psychological factors that plays a crucial role in the development of data literacy among students. That study showed that students with high levels of self-efficacy are better able to access, process, evaluate, and use data-based information as a basis for decision-making. Thus, the results of this study further strengthen the empirical evidence that the development of data literacy is influenced not only by academic ability but also by affective factors, specifically belief in one's own abilities.

On the other hand, the coefficient of determination (94.8%) in this study was higher than that found in several previous studies. This difference may be due to the relatively homogeneous characteristics of the study sample, the limited number of respondents, and the use of a single independent variable in the regression model. A sample drawn from a school setting with nearly identical characteristics can yield a stronger relationship among the components under study. Furthermore, the self-efficacy instrument developed based on Bandura's dimensions and the data literacy test tailored to the study's indicators allow for a more specific measurement of the relationship between the two variables.

The findings have implications for mathematics education, particularly regarding the importance of considering students' self-efficacy in data-based learning activities. However, this study did not experimentally test instructional strategies for increasing self-efficacy. Therefore, mastery experiences, constructive feedback, problem-based learning, group discussions, and contextual data tasks should be regarded as potential instructional recommendations rather than strategies proven to be effective by the present study. Future

experimental research could examine the effectiveness of these strategies in increasing self-efficacy and students' data literacy skills.

In addition, the findings of this study also contribute to the development of learning approaches that support the implementation of the Merdeka Curriculum. Data literacy is one of the key competencies in 21st-century learning because it is linked to critical thinking, decision-making, and problem-solving skills. Therefore, strengthening students' self-efficacy must be a priority in instructional planning so that students not only master mathematical concepts but also have the confidence to apply that knowledge in various real-world situations.

This study has several limitations that should be considered when interpreting the findings. First, the study employed an *ex post facto* design; therefore, the relationship between self-efficacy and data literacy skills cannot be interpreted as causal. Second, the sample consisted of only 30 students, all of whom came from one school, limiting the generalizability of the findings. Third, the study included self-efficacy as the only predictor, whereas data literacy may also be associated with other factors such as numeracy skills, critical thinking, learning motivation, learning experiences, and the learning environment. Fourth, the exceptionally high  $R^2$  value requires cautious interpretation and should be verified through research involving larger and more diverse samples. Therefore, future research is recommended to employ experimental or longitudinal designs, involve larger and more diverse samples, and consider multiple predictors.

This difference is likely due to the relatively homogeneous characteristics of the study sample, the limited number of respondents, and the fact that the study involved only one independent variable. These conditions may cause the relationship between variables to appear stronger than in studies involving larger and more diverse samples. Therefore, the results of this study should be interpreted with caution and should not be generalized to a broader population without further research.

Practically, the findings indicate the importance of considering self-efficacy in learning activities involving data literacy. However, strategies such as mastery experiences, constructive feedback, problem-based learning, group discussions, and data analysis exercises were not directly tested in this study. Therefore, these strategies may be considered

as recommendations for future research or instructional interventions rather than strategies proven by the present study to increase self-efficacy or data literacy skills.

## CONCLUSION

Based on the findings, self-efficacy showed a positive and statistically significant association with data literacy skills among the 30 eighth-grade students who participated in this study. The simple linear regression analysis produced a regression coefficient of 1.488 with a significance value of 0.001. The coefficient of determination was  $R^2 = 0.948$ , indicating that 94.8% of the variance in data literacy scores within the sample was statistically accounted for by self-efficacy.

However, because this study employed an ex post facto design without manipulating the self-efficacy variable, the findings should not be interpreted as evidence of a causal relationship. In addition, the sample consisted of only 30 students from one school, so the findings cannot be generalized to all junior high school students in Ulu Sosa Subdistrict. Further experimental or longitudinal research involving larger and more diverse samples is needed to examine whether changes in self-efficacy lead to changes in students' data literacy skills.

## ACKNOWLEDGMENTS

The author would like to thank Allah the Almighty for His mercy and blessings, which enabled this research to be successfully completed. The author also extends appreciation and gratitude to the academic advisor for providing guidance, feedback, and motivation throughout the research process. Gratitude is extended to the principal, teachers, and students of the junior high schools in Ulu Sosa Subdistrict for granting permission, providing assistance, and participating as respondents in this study. Furthermore, the author would like to thank his parents, family, and everyone who provided moral support, prayers, and assistance, making it possible for this research to be carried out and for this article to be successfully completed.

## REFERENCES

- Ananda, E. R., & Wandini, R. R. (2022). Analysis of students' mathematical literacy skills in relation to their self-efficacy. *Jurnal Obsesi: Journal of Early Childhood Education*, 6(5), 5113–5126.
- Arikunto, S. (2019). *Research Procedures: A Practical Approach*. Rineka Cipta.
- Azwar, S. (2021). *Reliability and Validity*. Pustaka Pelajar.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.
- Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Academy of Management Perspectives*, 26(4), 9–44.
- Fitriani, RN, & Pujiastuti, H. (2021). The influence of self-efficacy on mathematics learning outcomes. *Jurnal Cendekia: Journal of Mathematics Education*, 5 (3), 2793–2801
- Frailon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2023). *Preparing for life in a digital world: International computer and information literacy study*. Springer.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *\*International Statistical Review\**, 70(1), 1–25.
- Haditia, R., et al. (2021). Analysis of students' mathematical problem-solving skills. *\*Journal of Mathematics Education\**, 10(2), 123–130.
- Haqqul, A., et al. (2022). Students' mathematical literacy skills in solving data-based problems. *Journal of Education*, 11(1), 45–53.
- Indirwan, I., Suarni, W., & Priyatmo, D. (2021). The importance of self-efficacy for mathematics achievement. *Jurnal Sublimapsi*, 2(1), 61.
- Indrawati, F. A., & Wardono, W. (2019). The influence of self-efficacy on mathematical literacy and the development of 4C skills. In *PRISMA, Proceedings of the National Mathematics Seminar (Vol. 2, pp. 247–267)*.
- Ministry of Education and Culture. (2017). *Guidelines for the implementation of 21st-century skills*. Ministry of Education and Culture.
- Ministry of Education and Culture. (2020). *Education Policy in the Digital Age*. Ministry of Education and Culture.

- Khotimah, K., et al. (2020). The Effect of Self-Efficacy on Students' Mathematics Learning Outcomes. *Journal of Mathematics Education*, 9(2), 89–98.
- Kong, S. C., Wang, Y., & Dai, L. (2022). Data literacy and self-efficacy in secondary education. *Journal of Educational Technology & Society*, 25(2), 45–58.
- Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30–37.
- Marini, C. K., & Hamidah, S. (2014). The influence of self-efficacy, family environment, and school environment on the entrepreneurial interest of vocational high school students in the culinary arts program. *Journal of Vocational Education*, 4(2), 195–207.
- Narpila, SD (2016). Improving Students' Spatial Abilities and Self-Efficacy Through Inquiry-Based Learning Assisted by Cabri Software. *Jurnal Tarbiyah*, 23(1).
- Narpila, S.D. (2020). Efforts to Improve High School Students' Mathematical Self-Efficacy Through Inquiry-Based Learning. *Journal of Authentic Research in Mathematics Education (JARME)*, 2 (2), 108–118.
- National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. NCTM.
- OECD. (2021). *21st Century Readers: Developing Literacy Skills in a Digital World*. OECD Publishing.
- Priyastutik, S., et al. (2018). The Role of Mathematics in Everyday Life. *Journal of Mathematics Education*, 7(1), 1–10.
- Ridgway, J. (2016). Implications of the data revolution for statistics education. *\*International Statistical Review\**, 84(3), 528–549.
- Schild, M. (2004). Statistical literacy curriculum design. *Proceedings of the Section on Statistical Education*, 1–6.
- Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology*, 60, 101832.
- Sugiyono. (2019). *Educational research methods: Quantitative, qualitative, and R&D approaches*. Alfabeta.
- Usher, E. L., & Weidner, B. L. (2020). Social cognitive theory and motivation. *Journal of Educational Psychology*, 112(3), 580–595.
- Widoyoko, E. P. (2020). *Techniques for Developing Research Instruments*. Pustaka Pelajar.

- Wolff, A., Gooch, D., Caverio-Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *Journal of Community Informatics*, 12(3), 9–26.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91.