



Exploring the statistical literacy levels of eighth grade students: A gender-based analysis

Mohammad Auliya' Rahman¹, Arini Mayan Fa'ani^{2*}

^{1,2}Program Studi Tadris Matematika, UIN Maulana Malik Ibrahim Malang. Jalan Gajayana No.50
Malang

E-mail: iniarinimayan@uin-malang.ac.id *

Article received : June 28, 2025.

Article revised : December 2, 2025.

Article Accepted : December 28, 2025.

Article Publish : May 20, 2026

* Corresponding author

Abstract: Understanding statistical data is an important competency that students need to have in the information age. Statistical literacy encompasses not only the ability to read numbers, but also to interpret and connect data to real-life contexts. This study aims to describe the level of statistical literacy of eighth-grade students at SMP Negeri 13 Kota Malang in terms of gender, referring to Watson and Callingham's hierarchical model. This study employs a qualitative approach with a descriptive design. The subjects consist of one male student and one female student selected through purposive sampling. Data were collected through context-based essay questions and semi-structured interviews. The results indicate that male students are at the Informal level, tending to use imaginative narratives and personal experiences without systematically applying statistical concepts. Conversely, the female student was at the Consistent Non-Critical level, demonstrating the ability to consistently compare data using the concept of range, although she had not yet conducted critical analysis. These differences reflect variations in problem-solving approaches to statistical problems based on gender. This study recommends adaptive, contextual statistical learning that encourages students to think critically.

Keywords: statistical literacy; level statistical literacy; gender

INTRODUCTION

Statistics play an important role in modern life, which is dominated by information and data, so the ability to understand and use data appropriately has become an urgent need for every individual (Takaria and Rumahlatu (2016)). Statistical literacy is not just a technical skill of understanding numbers, but includes understanding graphs, tables, and data representations found in everyday life (Sušec et al., 2014). This ability is highly relevant in the context of mathematics education as it can improve the quality of students' decision-making based on statistical evidence (Gal, 2002).

However, various studies indicate that students' statistical literacy levels remain low and constitute one of the main challenges in mathematics education (Yulianti & Silvy, 2013). Many students are unable to interpret data deeply and are not accustomed to using more complex analysis strategies such as comparison or central tendency calculations (Rohmah, 2024). Even in the context of mathematics learning in schools, students tend to only be able to read single data points without being able to connect them to a broader context (Hafiyusholeh, 2015).

The literacy construct by Watson & Callingham (2003) is based on previous research by Watson (1997) using the Structure of the Observed Learning Outcome (SOLO) taxonomy from

Biggs & Collis (2004) in developmental psychology to categorize statistical literacy into a three-level hierarchy with increasing sophistication, namely, basic understanding of statistical terminology, understanding of statistical language and concepts when embedded in a broader social discussion context, and a questioning attitude to challenge claims made without proper statistical basis.

Watson & Callingham (2003) developed the three-level hierarchy into a construct of statistical literacy. This model is a six-level hierarchy that presents increasingly sophisticated thinking, ranging from idiosyncratic to critical mathematical. At the Idiosyncratic (Level 1) and Informal (Level 2) levels, students merely interact with the language and meaning of statistical terms. At the Inconsistent (Level 3) and Consistent Non-Critical (Level 4) levels, students begin to engage with the context and uncover the statistics embedded within that context. At the final two levels of this development, Critical (Level 5) and Critical Mathematical (Level 6), students can critically assess and challenge claims made in statistical reports and data.

Although there has been extensive research on statistical literacy, most of it still focuses on cognitive variables such as learning style (Syafiqah, 2022) or adversity quotient (Liyanti, 2024), and few have highlighted how gender factors play a role in shaping these abilities. However, in the field of education, gender differences are believed to influence learning styles, self-confidence levels, and how students approach mathematical problem-solving (Riwayani et al., 2024).

Social stereotypes that perceive males as superior in science and mathematics can negatively impact the academic performance of female students (Baiduri et al., 2023). For example, Santrock (2008) mentions that boys generally have an advantage in mathematical ability, as reflected in higher scores in several national assessments (Coley, 2001). This indicates that gender is not only a biological difference but also a social construct that influences how students learn and interact with course material (Kimmel, 2011).

In recent years, research on statistical literacy at the secondary school level has shown considerable growth. Various studies highlight that this ability is influenced by both cognitive and non-cognitive factors, ranging from learning styles (Syafiqah, 2022), cognitive styles (Suhana, 2020), to adversity quotient (Liyanti, 2024). In addition, research conducted by Kurnia et al. (2024) shows that the development of statistical literacy is gradual and highly dependent on students' ability to meaningfully interpret data within context. Nevertheless, most of these studies still tend to emphasize the quantification of learning outcomes, while qualitative investigations that uncover students' thinking processes directly through real contexts remain relatively limited.

Although statistical literacy has been widely studied, several research gaps are still apparent. At the junior high school level, for example, few studies specifically examine how gender differences influence the ways students read, interpret, and utilize statistical information. Yet, educational psychology literature suggests that gender-related characteristics often give rise to different reasoning patterns when solving problems. Moreover, the use of the Watson & Callingham model as an analytical framework to compare statistical literacy levels between male and female students is also rarely found, particularly

within Indonesian school contexts. The limited number of studies that integrate context-based tasks with in-depth interviews further reinforces the need to address this research gap.

This study aims to contribute to the development of statistical literacy research by providing an in-depth analysis of how male and female students think when solving context-based statistical literacy tasks. By employing the Watson & Callingham model, this study not only identifies the level of statistical literacy each participant demonstrates but also traces how they process data, formulate reasoning, and draw conclusions based on the contexts presented. The qualitative approach used through both within-case and cross-case analyses provides a more comprehensive depiction of each gender's reasoning patterns. Thus, this study is expected to enrich the literature on gender-based statistical literacy while serving as a reference for teachers in developing more adaptive and contextual statistics instruction.

Unfortunately, there is limited research specifically analyzing differences in statistical literacy levels between male and female students, particularly at the junior high school level. Initial observations at SMP Negeri 13 Kota Malang indicate that most students are unable to process and interpret data comprehensively, especially in drawing relevant conclusions based on the given statistical context. Therefore, this study aims to describe the statistical literacy levels of eighth-grade students at SMP Negeri 13 Kota Malang based on gender.

METHODS

This study uses a qualitative approach with a descriptive research design. This approach was chosen because the study aims to describe students' statistical literacy levels based on gender, as well as to gain a deeper understanding of how male and female students solve the statistical literacy questions provided. This research was conducted at SMP Negeri 13 Kota Malang with two eighth-grade students selected purposively as research subjects. The selection of subjects was based on the results of the initial statistical literacy test, taking into account gender representation and variations in ability to solve problems. The subjects were coded SL for male students and SP for female students.

Data collection in this study was conducted through two main instruments: tests and interviews. The statistical literacy test was developed based on the indicators in the statistical literacy level model according to [Watson & Callingham \(2003\)](#), which consists of six levels, ranging from Idiosyncratic to Critical Mathematical, The criteria can be seen in table 1.

The test was administered in the form of context-based essay questions to reveal students' abilities in understanding, interpreting, and analyzing statistical information in the form of graphs, tables, and narrative data. After the test was administered, semi-structured interviews were conducted with each subject to explore in greater depth the thoughts, reasons, and strategies used by students in completing the questions. The interviews were conducted individually and recorded for thematic analysis.

Table 1. Statistical Literacy Test Criteria

Level	Indicator
<i>Idiosyncratic</i>	<ul style="list-style-type: none"> - Students simply guess the answers without performing logical calculations or analysis. - Students calculate simply without understanding statistical concepts.
<i>Informal</i>	- Students perform basic one-step calculations on tables and graphs (such as addition and subtraction) based on observed values, but sometimes with imaginative stories.
<i>Inconsistent</i>	- Students draw conclusions but may not accompany them with appropriate statistical or mathematical justification.
<i>Consistent Non-Critical</i>	- Students demonstrate the appropriate application of statistical and mathematical concepts, including graphical characteristics.
<i>Critical</i>	- Students demonstrate qualitative interpretations and use of more complex mathematical or statistical concepts.
<i>Critical Mathematical</i>	- Students demonstrate advanced or critical statistical and mathematical skills related to concepts such as central tendency (mean, median, and mode) and measures of dispersion (range, variance, and standard deviation).

(source: [Kurnia et al., 2024](#))

The data obtained were analyzed using the analysis stages according to [Miles & Huberman \(2002\)](#), which consist of data reduction, data presentation, and conclusion drawing. Data reduction was carried out by selecting relevant data from the test and interview results, then classifying them based on gender and statistical literacy level categories. Next, the data was presented in a descriptive narrative form and quoted directly to reinforce the findings, so that the interpretation of the students' thinking processes could be seen concretely. The research conclusions were drawn by considering the patterns that emerged from each subject's data and relating them to the statistical literacy model and characteristics based on gender.

Data validity is maintained by using method triangulation techniques, which involve comparing test results with interview results to ensure that interpretations of students' abilities are more valid. Additionally, member checking is conducted with subjects to ensure that the researcher's interpretations align with the students' intentions and understanding. The entire data collection and analysis process is conducted systematically and carefully to ensure that the research results objectively reflect the actual conditions.

RESULTS AND DISCUSSION

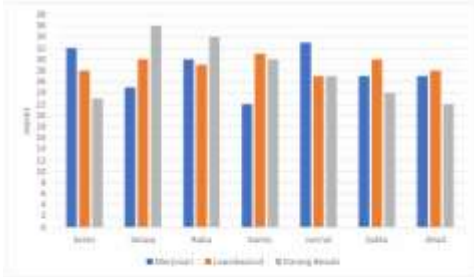
The data in this study were obtained from statistical literacy tests and semi-structured interviews with male and female students. The subjects were selected purposively. There

were two subjects analyzed, namely male students (SL) and female students (SP). Figure 1 below shows the statistical literacy questions used in the test.

Saal

Restoran Dapoer Kebuli milik Mas Ai memiliki tiga cabang di kota Malang, masing-masing dengan chef andalan yang memiliki kecepatan memasak berbeda. Suatu hari, seorang pelanggan bernama Mas Farhan datang dan meminta rekomendasi kepada Mas Ai mengenai cabang mana yang paling cepat untuk memproses pesanan dalam jumlah besar, karena Mas Farhan ingin memesan dengan jumlah yang banyak dalam waktu dekat.

Berikut adalah data rata-rata waktu memasak (dalam menit) untuk setiap 10 porsi nasi kebuli yang disajikan oleh masing-masing cabang selama tujuh hari terakhir berdasarkan Gambar 1 berikut.



Gambar 1 Grafik Waktu Memasak Setiap Cabang dalam 1 Minggu

Bantulah Mas Ai untuk menentukan cabang mana yang akan direkomendasikan kepada Mas Farhan! Tuliskan langkah-langkahnya dan jelaskan alasannya!

Questions

Mas Ai's Dapoer Kebuli restaurant has three branches in the city of Malang, each with a chef who has a different cooking speed. One day, a customer named Mas Farhan came and asked Mas Ai for a recommendation on which branch was the fastest at processing large orders, because Mas Farhan wanted to place a large order in the near future.

The following is the average cooking time (in minutes) for every 10 portions of nasi kebuli served by each branch over the past seven days, based on Figure 1 below.

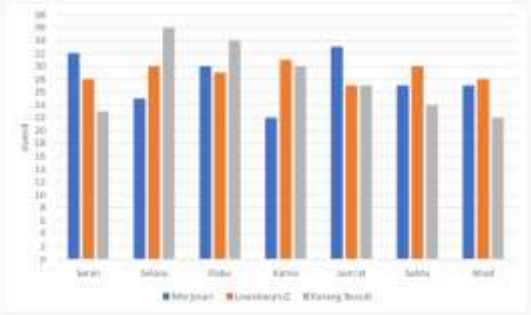


Figure 1 Cooking Time Chart for Each Branch in a Week

Help Mas Ai decide which branch to recommend to Mas Farhan! Write down your steps and explain your reasoning!

Figure 1 Statistical Literacy Questions

Male Students (SL)

Subject SL has completed the statistical literacy question given. SL determines the Dapoer Kebuli restaurant branch that is most appropriate to recommend to a customer named Mas Farhan, taking into account the daily cooking time at each branch. This can be seen in the results of SL's work in Table 2 below.

Table 2 SL article and interview no. 1

Carang Besuki. Waktu dijumlahkan waktu memasak untuk cabang Karang Besuki Senin = 23 menit Selasa = 36 menit Rabu = 34 menit Kamis = 30 menit Jumat = 27 menit Sabtu = 24 menit Minggu = 22 menit

If we add up the cooking time for the Karang Besuki branch, Monday = 23 minutes, Tuesday = 36 minutes, Wednesday = 34 minutes, Thursday = 30 minutes, Friday = 27 minutes, Saturday = 24 minutes, Sunday = 22 minutes.

P : You only added up the cooking time for the Karang Besuki branch, but did not do any other calculations. Can you explain why?

SL : The reason I only added up the cooking time at the Karang Besuki branch was because I wanted to show that I had seen the data at that branch and I found it easier to view and add up these figures.

Based on the results of Table 2, SL is aware of the addition operation, even though SL does not add the numbers, but only writes "if added." In interview number 1, SL stated that they performed addition of cooking times at the Karang Besuki branch, even though the mathematical operation was not explicitly visible in their answer sheet. SL explained that their

purpose was to demonstrate that they had observed and paid attention to the data available at that branch.

SL tends to only list the numbers from one branch (Karang Besuki) without performing explicit calculations or systematic comparisons with other branches. SL stated that they chose the addition method because it was considered “easier.” This statement indicates that SL is actually aware of other analysis methods that may be more appropriate but has not yet been able to apply them.

This is reflected in how SL writes his opinion about cooking times at the restaurant branches, focusing his attention on the Karang Besuki branch, particularly on Saturdays and Sundays. SL attempts to analyze his data, but his approach is still not sufficiently in-depth. As shown in Table 3 below.

Table 3 SL article and interview no. 2

<p>On Saturdays and Sundays, cooking time is only 24 and 22 minutes, respectively, which is the shortest compared to other branches. On Sundays, the Karang Besuki branch is the fastest at 22 minutes. Meanwhile, the Merjosari branch takes 27 minutes and the Lowokwaru branch takes 28 minutes.</p>
<p>P : Why did you write down the cooking time for Karang Besuki branches in detail in your answer?</p> <p>SL : I wrote down the cooking time for the Karang Besuki branch in detail because I think this branch has a faster cooking time, especially on weekends, which I think is important because people usually go to restaurants on weekends.</p>

Based on the results of Table 3, SL provided specific reasons for choosing the Karang Besuki branch. In interview number 2, SL's statement shows that SL conducted a simple analysis by identifying the advantages of the Karang Besuki branch based on shorter cooking times on weekends without performing statistical calculations. Instead, SL uses assumptions or personal experience to create an imaginative story that supports his analysis, as shown in interview number 2 SL: “especially on weekends, which I think is important because people usually go to restaurants on weekends.”

This statement is reinforced by SL's answer, which shows that SL uses assumptions or personal experience to create an imaginative story that supports his analysis, as shown in Table 4 below.

Table 4 SL article and interview no. 3

<p>People go to restaurants on weekends, so maybe Mr. Farhan will come to the restaurant on the weekend. So, I recommend the branch in Karang Besuki.</p>
<p>P : I see that you focus on cooking on Sundays. Why do you pay more attention to that day?</p> <p>SL : I paid more attention to Sundays because, in my experience, Sundays are usually busy days at restaurants. So, if Mr. Farhan is also likely to come on Sunday, I would recommend the branch with the fastest cooking time on Sundays, which is the Karang Besuki branch, at only 22 minutes.</p>

Based on the results of Table 4, SL paid special attention to the cooking time data on Sundays by simply observing the values displayed. SL identified that the Karang Besuki branch had the shortest cooking time on Sundays, which was 22 minutes, which he considered an advantage over other branches. Interestingly, SL creates an imaginative story based on his personal experience by stating that “Sundays are usually busy days at restaurants” and assuming that “Mr. Farhan will also likely come on Sundays.” This assumption is not based on statistical calculations but on SL's personal experience.

SL performed a basic one-step calculation by identifying the lowest cooking time on Sunday (22 minutes for the Karang Besuki branch) and simply comparing it with other branches. However, SL did not conduct more complex analyses such as calculating the average, median, or examining trends throughout the week. SL focused solely on one data point deemed important based on the imaginative story he created.

Based on this analysis, SL is at the informal statistical literacy level. At this level, SL analyzes values directly observed from the data (22 minutes cooking time on Sunday), performs simple calculations using addition operations, albeit implicitly, and uses assumptions or personal experience to create an imaginative story.

SL demonstrates a pattern of reasoning consistent with the Informal Level (Level 2) in the Watson & Callingham hierarchy. His responses focus only on surface-level data without looking at the overall pattern. When reading the restaurant branch data, SL immediately points to one branch without comparing values between days or between branches. SL is indeed capable of reading basic numbers, but does not connect them to a broader context. And the interview results show that SL often makes personal assumptions, such as imagining which branches are busy without any data to support this. [Watson & Callingham \(2003\)](#) state that Level 2 students tend to use intuition and everyday knowledge rather than statistical concepts. This characteristic is also in line with [Santrock \(2008\)](#) explanation that male students often make quick and less systematic decisions. Overall, SL has not demonstrated the ability

to compare values, understand variability, or provide data-based justifications, so his position at Level 2 is justified.

Female Students (SP)

In solving the statistical literacy problem given, SP determines the Dapoer Kebuli restaurant branch that is most appropriate to recommend to a customer named Mr Farhan, taking into account the difference in cooking time at each branch. This can be seen in the results of SP's work in Table 5 below.

Table 5 SP article and interview no. 1

<p>Dan waktu tercepat pada cabang lowokwaru adalah 27 menit sedangkan waktu terlama 31 menit. Jadi selisihnya adalah 4 menit. Jika di cabang merjosari waktu tercepat 22 menit dan waktu terlama 33 menit dan selisih 11 menit. Sedangkan di cabang karang besuki waktu tercepat 22 menit dan waktu terlama 36 menit dan selisihnya 14 menit.</p>
<p>And the fastest time in the Lowokwaru branch is 27 minutes. The longest time is 31 minutes, so the difference is 4 minutes. At the Merjosari branch, the fastest time is 22 minutes and the longest time is 33 minutes, with a difference of 11 minutes. At the Karang Besuki branch, the fastest time is 22 minutes and the longest time is 36 minutes, with a difference of 14 minutes.</p>
<p>P : Why are you paying more attention to the shortest and longest cooking times at each branch?</p> <p>SP : This difference indicates how stable the cooking time is. The smaller the difference, the more stable the cooking time. I think that stability is important because customers like Mr. Farhan may prefer a predictable waiting time rather than one that is unclear in terms of how long it will take.</p>

Based on the results of Table 5, SP uses the concept of range implicitly without realizing that what he is doing is applying a statistical concept. SP identifies that a smaller time difference indicates better stability in cooking time. In interview number 1 with SP, "This difference shows whether the cooking time is stable or not; the smaller the difference, the more stable the cooking time," indicating that SP understands the meaning of the calculations they perform. SP is able to read and compare cooking time data across branches by calculating the longest and shortest time differences for each branch. Indirectly, SP calculates the range, albeit implicitly. This difference calculation shows that SP is not only comparing numbers directly but also understands the importance of cooking time stability across each restaurant branch, as evident in SP's work results in the following Table 6.

Based on the results of Table 6, SP implicitly uses the concept of range to measure cooking time stability. In his answer, SP calculates the difference between the shortest and longest times in each branch, which is actually a range calculation. In interview number 2 with SP, SP explained their thought process by stating, "I concluded it by looking at and comparing the cooking times in that branch over the course of a week." This statement indicates that SP directly compared the available data. SP also mentioned that "the cooking time ranged from 27 to 31 minutes, with a difference of only 4 minutes," indicating that SP understood the concept of difference or range as a measure of stability, albeit implicitly.

Table 6 SP article and interview no. 2

<p>dan saya lihat dari grafik waktunya tidak terlalu berubah-ubah, jadi mas Farhan tidak perlu khawatir karena harus menunggu lama, dan jika di lihat dari waktu memasak setiap hari di cabang Lowokwaru</p> <ul style="list-style-type: none"> - hari senin : 28 menit - hari Selasa : 30 menit - hari Rabu : 29 menit - hari Kamis : 31 menit - hari Jumat : 27 menit - hari Sabtu : 30 menit - hari Minggu : 28 menit
<p>And I see from the graph that the time does not change much, so Mr. Farhan does not need to worry about having to wait a long time. Looking at the cooking time each day at the Lowokwaru branch: Monday = 28 minutes, Tuesday = 30 minutes, Wednesday = 29 minutes, Thursday = 31 minutes, Friday = 27 minutes, Saturday = 30 minutes, Sunday = 28 minutes.</p>
<p>P : You mentioned that cooking times at the Lowokwaru branch “don't vary too much.” How did you come to that conclusion?</p> <p>SP : I concluded this by looking at and comparing the cooking times at that branch over the course of a week. At the Lowokwaru branch, the cooking time ranged from 27 to 31 minutes, with a difference of only 4 minutes. when compared to the Merjosari branch, which has an 11-minute difference (22–33 minutes), and the Karang Besuki branch with a 14-minute difference (22–36 minutes), it is clear that the cooking time at the Lowokwaru branch is significantly shorter. This indicates a more consistent cooking time at that branch.</p> <p>P : Why did you write down the cooking time for the Lowokwaru branch in such detail?</p> <p>SP : I wrote down the cooking times for the Lowokwaru branch in detail to show how the times at this branch are consistent from day to day. Monday 28 minutes, Tuesday 30 minutes, Wednesday 29 minutes, and so on. This indicates that the time is not significantly longer compared to other branches, which is why I chose the Lowokwaru branch because its cooking time is stable, which I believe is a very important factor to recommend to Mr. Farhan.</p>

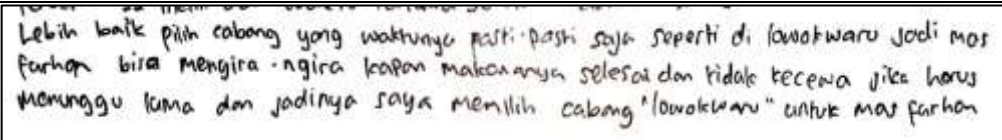
SP also emphasized the importance of cooking time stability by stating, “I wrote down the cooking times for the Lowokwaru branch in detail to show how the times at this branch are stable from day to day.” SP listed the cooking times for each day in sequence (Monday 28 minutes, Tuesday 30 minutes, Wednesday 29 minutes, and so on) to support his argument about the stability of cooking time at the Lowokwaru branch. This shows that SP can read and compare data well, identify patterns of stability in the data, and use the results of the difference calculation (range) to make recommendations, but SP does not realize that what he is doing is a range.

SP interprets the data in a real-world context, considering that cooking time stability may be more important to customers than cooking times that are sometimes very fast but inconsistent. This can be seen in the results of SP's work in Table 7.

Based on the results of Table 7, SP demonstrated the ability to consistently use statistical calculations to compare the three restaurant branches. SP applied the concept of range (the difference between the maximum and minimum values) as a measure of cooking time variability in each branch, although he did not explicitly mention statistical terms. The

statement “he can estimate the waiting time to be around 27-31 minutes” indicates that SP understands the impact of small time differences on restaurant service. SP consistently uses range calculations as the basis for making recommendations, indicating that SP understands that a small variability measure indicates better consistency in cooking time. Although SP does not use formal statistical terms such as range, SP has applied the concept appropriately in comparing the three restaurant branches. Additionally, SP does not calculate measures of central tendency such as the mean, but consistently uses the range to assess the stability of cooking times across branches. However, SP has not demonstrated critical analysis of the data, such as questioning the reliability of the data provided.

Table 7 SP article and interview no. 3


<p>It's better to choose a branch with fixed opening hours, such as the one in Lowokwaru. That way, Farhan can estimate when his food will be ready and won't be disappointed if he has to wait a long time. So, I chose the Lowokwaru branch for Farhan.</p>
<p>P : How would you explain your statement, “It's better to choose a branch with a definite schedule”?</p> <p>SP : What I mean is that customers can know approximately how long they have to wait. At the Lowokwaru branch, the cooking time does not vary much from day to day, with only a 4-minute difference between the fastest and slowest. So if Mr. Farhan goes there, he can estimate that the waiting time will be around 27-31 minutes.</p>

Based on this analysis, it can be concluded that SP is at the Consistent Non-Critical level of statistical literacy. At this level, SP is able to consistently perform statistical calculations (range) to analyze data and determine answers. SP can also interpret the results of their calculations in a real-world context and use them as a basis for making recommendations. However, SP has not demonstrated the ability to perform critical analysis of the data.

Unlike SL, SP demonstrates reasoning consistent with the Consistent Non-Critical Level (Level 4). SP reads data thoroughly, compares values between days, and observes stability patterns. The decisions he makes are always supported by clear reasoning and refer to specific numbers, not assumptions. And the interview, SP explains his choices by showing which branches are the most consistent and which are fluctuating. SP demonstrates an understanding of context and is able to relate it to numerical information, a key feature of Level 4. Research on gender in mathematics education (Baiduri et al., 2023; Coley, 2001) shows that female students tend to be meticulous and consider details, which is evident in SP's approach. Although he has not yet reached Levels 5–6, which are critical in nature, his explanations show a stable and evidence-based pattern of reasoning, making Level 4 the most appropriate classification.

The comparative analysis revealed several important differences between SL and SP. First, in interpreting data, SL only reads numbers separately without looking at the data series. In contrast, SP examines the entire dataset and can identify patterns of stability between branches. Second, in terms of argumentation style, SL often uses personal assumptions, while SP consistently provides number-based reasoning. SP cites specific values to explain his choices, while SL does not provide strong justification. Third, in terms of understanding statistical concepts, SP demonstrates awareness of data variability and patterns, while SL does not demonstrate such understanding. SP understands that data should not only be viewed in terms of the highest values, but also in terms of patterns and stability. Fourth, these differences are also relevant to gender theory. [Santrock \(2008\)](#) explains that male students tend to make quick decisions based on intuition, while studies such as [Coley \(2001\)](#) and [Baiduri et al. \(2023\)](#) find that female students are usually more analytical and cautious. This pattern is clearly seen in the cases of SL and SP.

The findings of this study reveal clear differences in how male and female students interpret and use statistical information when responding to contextual problems. These differences reflect the unique ways in which students reason with data, consistent with earlier research showing that statistical literacy develops differently across grade levels and individual characteristics ([Kurnia et al., 2024](#)).

The male student relied heavily on intuition and everyday experience. Instead of examining the data systematically, he drew conclusions based on assumptions such as the belief that Sundays are typically busier to justify his answer. This type of reasoning is frequently found among students whose statistical understanding is still emerging, as they tend to depend on familiar narratives or personal impressions rather than on the numerical information provided ([Hafiyusholeh, 2015](#)). Although the male student was able to identify the shortest cooking time, he did not compare the data in a structured manner. This aligns with findings from [Maryati & Priatna \(2018\)](#), who noted that students with lower levels of statistical literacy often resort to story-based explanations when unsure how to interpret numerical data. Such tendencies place him at the Informal level in Watson and Callingham's hierarchy, where students begin connecting data to context but have not yet developed consistent statistical reasoning. Similar surface-level reasoning patterns have also been reported among Indonesian students in other contexts ([Negara et al., 2019](#)).

In contrast, the female student demonstrated a more systematic and data centered approach. She compared the longest and shortest cooking times, examined the stability of the data, and used these observations to justify her conclusion. Her focus on consistency reflects a more advanced level of thinking. [Koga \(2022\)](#) explains that students at higher stages of statistical literacy tend to interpret variability more thoughtfully and use it meaningfully in decision-making. The female student's careful numerical comparison and her ability to link the data directly to the contextual problem indicate that she is operating at the Consistent Non-Critical level. This finding is also supported by studies showing that female students often engage more deeply with numerical patterns and contextual details ([Riwayani et al., 2024](#)).

These contrasting approaches suggest that statistical literacy is shaped not only by students understanding of mathematical concepts but also by how they think about problems and connect numerical information with real situations. Curiosity and the ability to interpret data meaningfully also play an important role in this process (Setiani & Suyitno, 2021). Thus, the differences observed in this study should not be taken to imply that one gender performs better than the other, but rather that each student demonstrates a different reasoning pathway shaped by their cognitive tendencies and learning experiences.

The findings carry several implications for the design of statistical learning in junior high school settings. First, male students like SL, who are positioned at the Informal level, often need structured scaffolding that gradually moves them from intuitive, narrative-based thinking toward more systematic analysis. Activities that encourage students to compare data across categories, justify their choices numerically, and reflect on the difference between intuitive assumptions and data-based conclusions can help strengthen their statistical reasoning.

Second, female students like SP, who operate at the Consistent Non-Critical level, show strong potential for more advanced reasoning but still require opportunities to evaluate the credibility and reliability of the data. Tasks that ask students to question assumptions, consider sample size, compare multiple statistical measures, or analyze the limitations of the data can help develop higher-level statistical literacy.

Third, the contrast between SL and SP highlights the need for adaptive teaching approaches. Students who rely on quick, intuitive decisions benefit from tasks that slow down the reasoning process and guide them through step by step comparison. Meanwhile, students who naturally pay attention to details may need challenges that encourage them to examine broader contexts and scrutinize the strength of data based claims.

Overall, these findings underline the importance of contextual, discussion-based instruction that emphasizes argumentation, justification, and critical evaluation rather than focusing solely on procedural calculations. Such approaches may support students progression through the statistical literacy hierarchy toward more critical and mathematically sound reasoning. Teachers can facilitate this progression by creating opportunities for classroom discussions that revolve around real data, helping students become more comfortable interpreting tables and graphs, and guiding them to question the meaning and reliability of the data they encounter.

The hierarchical model proposed by Watson & Callingham (2003) proves effective in capturing students' statistical literacy in a more holistic way. It does not merely assess whether students' answers are correct, but also examines the depth of their reasoning, the strength of their arguments, and the way they connect data with context. By using this model, teachers can better understand how students think and can offer feedback that supports their development toward higher levels of literacy.

Finally, the results of this study align with earlier findings by Andriatna & Kurniawati (2021), who emphasized the role of conceptual understanding in statistical learning, as well as research by Syafiqah (2022) and Suhana (2020) which showed that learning styles and

cognitive styles influence how students approach and interpret statistical tasks. Taken together, these studies reinforce the idea that statistical literacy is a complex ability involving cognitive, affective, and contextual dimensions. Therefore, the differences observed between male and female students in this study represent variations in learning tendencies rather than differences in capability. These findings highlight the need for adaptive teaching strategies and context-rich questions that encourage students to think critically, analytically, and reflectively about data.

CONCLUSION

This study aims to describe the statistical literacy level of eighth-grade students at SMP Negeri 13 Malang City in terms of gender, referring to Watson and Callingham's hierarchical model. Based on the analysis results, it was found that male and female students showed variations in achieving statistical literacy levels, both in terms of understanding data, constructing arguments, and drawing conclusions based on context. However, these differences do not fully reflect the direct influence of gender but are more influenced by thinking patterns, problem-solving strategies, and learning experiences that shape students' ability to interpret and use statistical data contextually.

The scientific justification for these results indicates that statistical literacy is a complex competency that requires active engagement in contextual, meaningful learning that encourages students to think critically. The application of the Watson and Callingham model has proven helpful in identifying and classifying students' ability levels in a more structured manner, enabling teachers to develop adaptive learning strategies. This model also opens up space to examine students' thinking processes more deeply, not merely evaluating the correctness or incorrectness of answers, but also the logic and relevance of the arguments students construct regarding the data.

The application of these findings points to the need for innovation in statistics education at the junior high school level, with a data-driven approach and tasks that emphasize interpretation and argumentation. This study also suggests that future experiments should include a broader population and examine the influence of other variables such as learning style, motivation, or environmental factors on the development of statistical literacy. Longitudinal research could also be conducted to evaluate the sustainability of the learning approach's influence on students' statistical abilities over time. Thus, the results of this study not only expand theoretical insights into gender-based statistical literacy but also provide practical contributions to improving the quality of mathematics education in schools.

Despite providing meaningful insights, this study has several limitations. The analysis was conducted on only two students, which restricts the generalizability of the findings. The statistical tasks used were limited to a specific contextual format, so they did not fully represent the broader scope of statistical literacy. In addition, the interview duration may not have captured all possible reasoning pathways that students could demonstrate. Future research should involve a larger and more diverse sample to explore gender-related patterns more comprehensively. It will also be valuable to include a wider range of statistical tasks,

such as inference, probability, and evaluation of real-world claims, to obtain a more complete picture of students' statistical literacy. Longitudinal studies are also recommended to monitor how students progress across the Watson and Callingham levels over time and how learning interventions influence their reasoning development.

REFERENCES

- Andriatna, R., & Kurniawati, I. (2021). Analisis Level Literasi Statistik Mahasiswa Calon Guru Matematika. *Transformasi : Jurnal Pendidikan Matematika Dan Matematika*, 5(2), 619–632. <https://doi.org/10.36526/tr.v5i2.1497>
- Baiduri, I., Hasanah, N., Maulana, F., & Anshori, M. I. (2023). Gender dan Kepemimpinan : Sebuah Kajian Literatur. *Jurnal Ilmu Manajemen, Ekonomi Dan Kewirausahaan*, 3(2), 179–204. <https://doi.org/10.55606/jimek.v3i2.1782>
- Biggs, J. B., & Collis, K. F. (2004). *SOLO Taxonomy and Assessing Learning to Learn*.
- Coley, R. J. (2001). Differences in the Gender Gap : Comparisons Across Racial / Ethnic Groups in Education and Work. *Policy Information Report*, 1–51.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1–25. <https://doi.org/10.1111/j.1751-5823.2002.tb00336.x>
- Hafiyusholeh, M. (2015). Literasi statistik dan urgensinya bagi siswa. *Wahana*, 64(1), 1–8. <https://doi.org/https://jurnal.unipasby.ac.id/index.php/whn/article/view/531/390>
- Kimmel, M. (2011). Explanations of Gender. In *The Gendered Society* (pp. 19–111).
- Koga, S. (2022). Characteristics of statistical literacy skills from the perspective of critical thinking. *Teaching Statistics*, 44(2), 59-67. <https://doi.org/10.1111/test.12302>
- Kurnia, A. B., Lowrie, T., & Patahuddin, S. M. (2024). The development of high school students' statistical literacy across grade level. *Mathematics Education Research Journal*, 36(Suppl 1), 7–35. <https://doi.org/10.1007/s13394-023-00449-x>
- Liyanti, R. P. (2024). Analisis Literasi Statistik Ditinjau Dari Adversity Quotient Siswa.
- Maryati, I., & Priatna, N. (2018). Analisis Kemampuan Literasi Statistik Siswa Madrasah Tsanawiyah dalam Materi Statistika. *Journal of Medives : Journal of Mathematics Education IKIP Veteran Semarang*, 2(2), 205. <https://doi.org/10.31331/medives.v2i2.640>
- Miles, M. B., & Huberman, A. M. (2002). The qualitative researcher's companion. *Sustainability (Switzerland)*, 11(ue 1).

- Negara, H. R. P., Santosa, F. H., & Ibrahim, M. (2019). Mengidentifikasi dan Mengembangkan Kemampuan Penalaran Statistis. *MES: Journal of Mathematics Education and Science*, 5(1), 29–41. <https://jurnal.uisu.ac.id/index.php/mesuisu>
- Riwayani, I., E. S., Perdana, R., & Soeharto. (2024). Analyzing students' statistical literacy skills based on gender, grade, and educational field. *International Journal of Evaluation and Research in Education*, 13(2), 842–851. <https://doi.org/10.11591/ijere.v13i2.26299>
- Rohmah, N. N. (2024). Lansia Menyikapi Misinformasi: Bentuk, Strategi dan Upaya Penanggulangan Hoaks dalam Konteks Pemilu 2024 di Nusa Tenggara Barat. *Journal of Islamic Communication Studies (JICoS)*, 2(1), 20–32. <https://doi.org/10.15642/jicos.2024.2.1.20-32>
- Santrock, J. W. (2008). *Psikologi pendidikan (Edisi ke-2)*. Kencana Prenada Media Group.
- Setiani, N. W., & Suyitno, A. (2021). Kemampuan Membaca Data dan Rasa Ingin Tahu Siswa Terhadap Kemampuan Literasi Statistik. *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, 13(2), 257–270. <https://doi.org/10.37680/qalamuna.v13i2.915>
- Suhana. (2020). Analisis Literasi Statistik Siswa Madrasah Tsanawiyah Ditinjau Dari Gaya Kognitif.
- Sušec, M. P., Muravec, N. J., & Stančić, H. (2014). Statistical literacy as an aspect of media literacy. *Medijska Istrazivanja*, 20(2), 131–153.
- Syafiqah, N. F. (2022). Analisis Literasi Statistik Ditinjau Dari Gaya Belajar. *Journal of Economic Perspectives*, 2(1), 1–4.
- Takaria, J., & Rumahlatu, D. (2016). The Effectiveness of CPS-ALM Model in Enhancing Statistical Literacy Ability and Self Concept of Elementary School Student Teacher. *Journal of Education and Practice*, 7(25), 44–49. www.iiste.org
- Watson, J., & Callingham, R. (2003). Statistical Literacy: a Complex Hierarchical Construct. *Statistics Education Research Journal*, 2(2), 3–46. <https://doi.org/10.52041/serj.v2i2.553>
- Watson, J. M. (1997). Assessing Statistical Thinking Using the Media. In *The Assessment Challenge in Statistics Education* (pp. 107–121).
- Yulianti, N., & Silvy, M. (2013). Sikap Pengelola Keuangan Dan Perilaku Perencanaan Investasi Keluarga Di Surabaya. *Journal of Business and Banking*, 3(1), 57–68.