AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 15, No. 2, 2026, 1274-1290

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

LEARNING OBSTACLES IN DESCRIPTIVE STATISTICS: CHALLENGES IN STATISTICAL THINKING FOR PROSPECTIVE MATHEMATICS TEACHERS

Eliva Sukma Cipta[1*], Didi Suryadi[2], Tatang Herman[3], Al Jupri[4]

[1*] Universitas Islam Nusantara, Bandung, Indonesia
[2,3,4] Universitas Pendidikan Indonesia, Bandung, Indonesia
*Corresponding author: elivasukmacipta@uninus.ac.id

Received 18 November 2025; Revised 01 June 2026; Accepted 27 June 2026

Abstract

Statistical thinking is an essential competency for prospective mathematics teachers, as it enables them to understand, organize, represent, analyze, and interpret data effectively. However, many students still experience difficulties in understanding and applying descriptive statistics concepts, which may hinder the development of their statistical thinking skills. Therefore, identifying learning obstacles is necessary as a basis for designing more effective learning experiences. This study aims to identify the learning obstacles encountered by prospective mathematics teachers in descriptive statistics courses. The study employed Didactical Design Research (DDR) using a qualitative methodology within an interpretive paradigm. The participants were 38 prospective mathematics teachers enrolled at a private university in Bandung, Indonesia. The research instruments consisted of a statistical thinking test, semi-structured interview guidelines, and document analysis. Data were collected through tests, interviews, and document analysis and were analyzed through the processes of identification, classification, reduction, and verification. Data trustworthiness was established through methodological and source triangulation. The findings revealed that students experienced difficulties across all indicators of statistical thinking, including describing, organizing, representing, analyzing, and interpreting data. These difficulties were reflected in misconceptions regarding the relationship between mean, median, mode, and distribution skewness, the application of measures of position, the interpretation of data variability, and the analysis of relative performance. The study identified ontogenic, epistemological, and didactical obstacles that hindered the development of students' statistical thinking. These findings provide an empirical basis for developing a hypothetical learning trajectory to enhance statistical thinking in descriptive statistics courses.

Keywords: Descriptive statistics; learning obstacle; statistical thinking.

Abstrak

Kemampuan berpikir statistis merupakan kompetensi penting yang harus dimiliki oleh mahasiswa calon guru matematika. Namun, mahasiswa masih mengalami berbagai kesulitan dalam memahami dan menerapkan konsep statistika deskriptif, sehingga menghambat perkembangan kemampuan berpikir statistis. Oleh karena itu, identifikasi learning obstacle diperlukan sebagai dasar dalam merancang pembelajaran yang lebih efektif. Penelitian ini bertujuan untuk mengidentifikasi learning obstacle mahasiswa calon guru matematika dalam perkuliahan statistika deskriptif. Penelitian ini menggunakan pendekatan Didactical Design Research (DDR) dengan metodologi penelitian kualitatif dalam paradigma interpretif. Partisipan penelitian terdiri atas 38 mahasiswa calon guru matematika di salah satu perguruan tinggi swasta di Bandung. Instrumen penelitian meliputi tes kemampuan berpikir statistis, pedoman wawancara semi-terstruktur, dan studi dokumentasi. Data dikumpulkan melalui tes, wawancara, dan analisis dokumen, kemudian dianalisis melalui tahapan identifikasi, klasifikasi, reduksi, dan verifikasi data. Keabsahan data diperoleh melalui triangulasi metode dan sumber. Hasil penelitian menunjukkan bahwa mahasiswa mengalami kesulitan pada seluruh indikator berpikir statistis, yaitu mendeskripsikan, mengorganisasikan, merepresentasikan, serta menganalisis dan menginterpretasikan data. Kesulitan tersebut ditunjukkan oleh miskonsepsi dalam menghubungkan mean, median, dan modus dengan kemiringan distribusi, penggunaan ukuran letak data, interpretasi variabilitas data, serta analisis performansi relatif. Temuan penelitian mengidentifikasi adanya hambatan ontogenik, epistemologis, dan didaktis. Temuan ini memberikan dasar empiris bagi penyusunan hypothetical learning trajectory untuk mengembangkan kemampuan berpikir statistis mahasiswa.

***Kata kunci**: Berpikir statistis; hambatan belajar; statistika deskriptif.*

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

INTRODUCTION

Descriptive statistics is one of the essential topics in mathematics education at the higher education level, particularly in mathematics education programs, because it supports students' ability to use numerical summaries and data visualizations to describe and communicate data meaningfully (Setiawan & Sukoco, 2021). Mastery of this topic also forms the foundation for statistical thinking, which encompasses the skills of identifying, representing, analyzing, and interpreting data across various contexts (Friedrich et al., 2024; Cipta et al., 2026). These competencies remain fundamental in contemporary statistics education and are increasingly recognized as essential for promoting statistical literacy, data interpretation, and data-informed decision making in the twenty-first century (Engel, 2017; Schreiter et al., 2024).

A growing body of research reports that mathematics education students continue to experience substantial difficulties in understanding fundamental statistical concepts as well as their applications in problem-solving situations. Azis & Dahlan, (2024) reported that mathematics education students demonstrated low statistical literacy, particularly in understanding fundamental statistical concepts, interpreting data, and communicating statistical information. Şahal & Koklu (2026) found that pre-service mathematics teachers experienced difficulties in conducting statistical investigations, particularly in formulating investigative questions, identifying relevant variables, selecting representative samples, and planning data collection and analysis.

These challenges also appear in students' conceptual and procedural engagement with descriptive statistics. Rohimah et al. (2026) documented conceptual and procedural errors among pre-service teachers, including inaccuracies in applying descriptive statistics formulas and constructing statistical representations such as frequency tables, histograms, and pie charts. Furthermore, Setiawan & Sukoco (2021) found that undergraduate students experienced difficulties in selecting appropriate descriptive statistics and meaningful graphical representations, with many students being unfamiliar with histograms, boxplots, stem-and-leaf plots, and dot plots for representing continuous data.

International findings show similar patterns, particularly students' miscon-ceptions related to measures of central tendency and variability (Boels et al., 2019; Chaphalkar & Wu, 2020); Sanchez, 2023) as well as their limited ability to connect data with contextual meanings during statistical reasoning processes (Frischemeier & Schnell, 2023); Oslington et al., 2024).

One of the underlying causes of these difficulties is the presence of learning obstacles experienced by students throughout the learning process. Brousseau, (2002) classifies learning obstacles into three categories: epistemological obstacles, which relate to the structure and nature of the concepts; ontogenic obstacles, which stem from students' individual learning experiences; and didactical obstacles, which arise from instructional design or teaching practices. Mapping these obstacles is essential to ensure that instruction is aligned with students' needs and to minimize the emergence of conceptual errors (Suryadi, 2019; Artigue, 2009).

In the learning of descriptive statistics, students are expected not only to perform calculations but also to

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

comprehend the meaning behind the results and use the information to make informed decisions. Statistical thinking has thus become an essential competence, particularly in the era of big data (Ben-Zvi & Makar, 2016; Prodromou & Dunne, 2017; Cipta et al., 2024). Consequently, instructors must understand the various forms of learning obstacles that emerge during the learning process so that instruction does not focus solely on procedures but also supports the meaningful development of students' conceptual understanding (Cipta, Suryadi, Herman, Jupri, et al., 2024).

Although numerous studies have examined students' difficulties in learning statistics, most of these investigations remain general in nature and have not systematically mapped learning obstacles based on indicators of statistical thinking. Previous research has also tended to examine students' difficulties at particular stages of statistical investigations, such as formulating statistical questions, identifying relevant variables, selecting representative samples, and planning data collection and analysis, rather than linking these difficulties comprehensively to epistemological, ontogenic, and didactical factors (Hidayat et al., 2023; Şahal & Koklu, 2026). Accordingly, the literature remains limited in providing a structured identification of learning obstacles in descriptive statistics through a statistical-thinking framework.

Building on this research gap, the present study aims to identify the types of learning obstacles encountered by students when solving descriptive statistics problems, based on four indicators of statistical thinking. The findings of this study are expected to contribute to the development of more effective didactical designs for teaching descriptive statistics by drawing on the theory of didactical situations (Brousseau, 2002). In addition, the results may assist instructors in designing instruction that better facilitates students in constructing a comprehensive and meaningful understanding of statistical concepts.

METHODS

This study employed a qualitative approach with a descriptive method. This approach was selected because the research aims to identify the learning obstacles experienced by prospective mathematics teachers in understanding descriptive statistics and to examine the epistemological, ontogenic, and didactical factors underlying these obstacles. The research design refers to the framework of Didactical Design Research (DDR) developed by Cipta et al. (2024) and Suryadi (2019), however, the present study was conducted only up to the stage of analyzing the didactical situation and did not proceed to the development of a subsequent didactical design.

The study was conducted during the odd semester of the 2024/2025 academic year at a private university in Bandung. The participants consisted of 35 students from the Mathematics Education Study Program who had completed the Basic Statistics course. Participants were selected purposively based on the following criteria: (1) they had previously received instruction on descriptive statistics, and (2) the course instructor agreed to collaborate during the research process (Sugiyono, 2022).

The primary research instrument was a set of four open-ended questions constructed based on four indicators of statistical thinking: (1) describing data, (2) organizing data, (3) representing data, and (4) analyzing and interpreting data. Each indicator was represented by

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

one item requiring students to work with real data, graphs, or tables. Test validation was conducted through expert judgment by three specialists in statistics and mathematics education. Revisions to the instrument were made based on expert feedback regarding indicator alignment, clarity of context, and cognitive demand.

In addition to the written test, supporting instruments included semi-structured interview guides for both students and the course instructor, as well as documentation sheets for reviewing the syllabus, instructional materials, and assessment tasks previously used in the course. Students selected for interviews were chosen purposively based on variation in test performance (high, medium, low) to explore their thinking processes, problem-solving strategies, and the specific difficulties they experienced.

The research was conducted through several stages. First, students completed the statistical thinking test during a 90-minute session. Second, test responses were analyzed using a scoring rubric that assessed procedural accuracy, quality of representation, accuracy of interpretation, and contextual relevance. Third, in-depth interviews were conducted with nine selected students to obtain a more comprehensive understanding of the sources of difficulty. Fourth, an interview with the course instructor was carried out to investigate the instructional context. Fifth, a document analysis was conducted to examine the alignment between instructional materials, examples, and exercises with the students' observed difficulties.

Data were analyzed through a systematic thematic analysis involving data familiarization, selection and reduction of relevant data segments, coding, categorization of related codes, and theme development to identify patterns of students' learning obstacles (Naeem et al., 2024). Data from tests, interviews, and documents were cross-checked using methodological triangulation and source triangulation to enhance the credibility and consistency of the findings (Patton, 2022; Moleong, 2007).

Following the thematic analysis, the identified difficulties were further classified into three categories of learning obstacles based on the framework proposed by Brousseau, (2002) and adapted within Didactical Design Research (Suryadi, 2019). Epistemological obstacles were identified when students' difficulties stemmed from limited conceptual understanding or misconceptions regarding statistical concepts and their application in different contexts. Ontogenic obstacles were identified when difficulties were associated with students' cognitive readiness, prior knowledge, or learning experiences that hindered their ability to engage with the tasks. Didactical obstacles were identified when difficulties were related to instructional factors, including teaching approaches, learning materials, classroom inte-ractions, or assessment practices. The categorization process was conducted by comparing evidence obtained from students' written responses, interview transcripts, instructor interviews, and document analysis. The consistency of the classifications was verified through triangulation across data sources.

RESULTS AND DISCUSSION

For the first statistical thinking indicator, namely describing data, students were asked to determine whether the data distribution was symmetrical or not by examining the skewness of the provided histogram and

AKSIOMA: Jurnal Program Studi Pendidikan Matematika ISSN 2089-8703 (Print)
Volume 15, No. 2, 2026, 1274-1290 ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

by comparing the mean, median, and mode. The analysis of students' responses revealed a variety of difficulties in understanding these concepts. Several examples of student responses are presented in Figure 1.

Based on the results of the written test, it was found that some students recalculated the mean even though the histogram already provided its value, as shown in Figure 1. Students were unable to establish the relationship between the mean, median, mode, and the skewness of the data distribution displayed in the histogram. This finding was reinforced by interview results indicating that students performed the recalculation simply because the task asked them to compare the values, without realizing that the relevant information had already been presented in the histogram. When further asked about the connection between the histogram and the measures of central tendency, most students stated that they did not know or had forgotten.

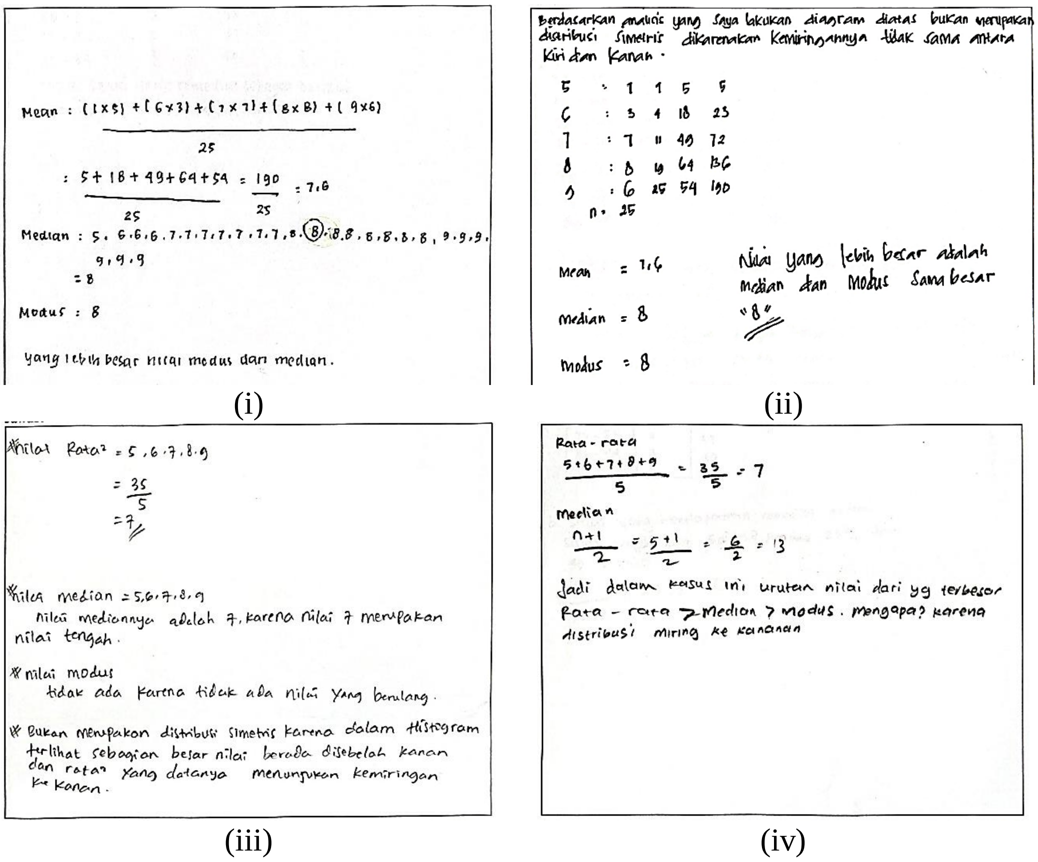


Figure 1. Student Answers to Data Describing Indicators

In addition, errors were also found in students' understanding of the concepts of the median position and the median value. Many students determined the value of the median using only the formula for the median's position, without proceeding to identify the actual median value based on the data distribution. A similar mistake appeared in students' treatment of the mean, where several of them calculated the mean solely from the number of data variations rather than considering the total number of data points. This can be seen in Figures 1(iii) and 1(iv). In determining the mode, students also

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

misunderstood the nature of the data, assuming that the dataset consisted of single data points with a frequency of one, whereas the histogram clearly provides frequency values that should be taken into account.

These difficulties were then categorized into several types of learning obstacles. The first is an epistemological learning obstacle, which arises from students' limited understanding of fundamental concepts in descriptive statistics. Students exhibited misconceptions in connecting the mean, median, and mode with the skewness of a distribution. For instance, they recalculated the mean even though it was already provided, and they misunderstood how these three measures of central tendency relate to symmetric and asymmetric distributions. Misconceptions also appeared in students' interpretation of the median, where the position of the median was mistaken for its actual value, as well as in identifying the mode, where errors occurred because students treated the data as if each value had a frequency of one.

Second, an ontogenic learning obstacle was identified, which relates to students' cognitive development and prior learning experiences. Most students had not yet reached a stage where they could grasp more complex statistical concepts, particularly those requiring the integration of foundational ideas with visual interpretations, such as recognizing the skewness of a histogram. This difficulty is reinforced by the limited learning experiences students had prior to the course, as they were rarely exposed to tasks that required graphical interpretation or analysis of the relationships among measures of central tendency.

Third, a didactical learning obstacle also emerged, referring to barriers arising from instructional processes, teaching methods, learning materials, and classroom interactions. Students tended not to receive sufficient practice or learning experiences that emphasized practical application or graphical interpretation of statistical concepts. The instructional materials used were largely procedural, providing formulas and steps without explicitly connecting them to data visualizations or real-world contexts. In addition, interactions between lecturers and students, as well as the feedback provided, were perceived as insuffi-ciently specific to help students address and correct their misconceptions.

These findings indicate that students' understanding of descriptive statistical concepts remains limited, particularly in describing data based on distribution skewness and the relationships among measures of central tendency. As noted by Boels et al., (2019) statistical thinking requires a solid conceptual foundation as well as diverse learning experiences both in numerical computation and in the interpretation of data visualizations. Therefore, the results of this study provide important insights for the development of didactical designs that place greater emphasis on conceptual strengthening, visual interpretation, and contextual exercises that can help reduce students' learning obstacles.

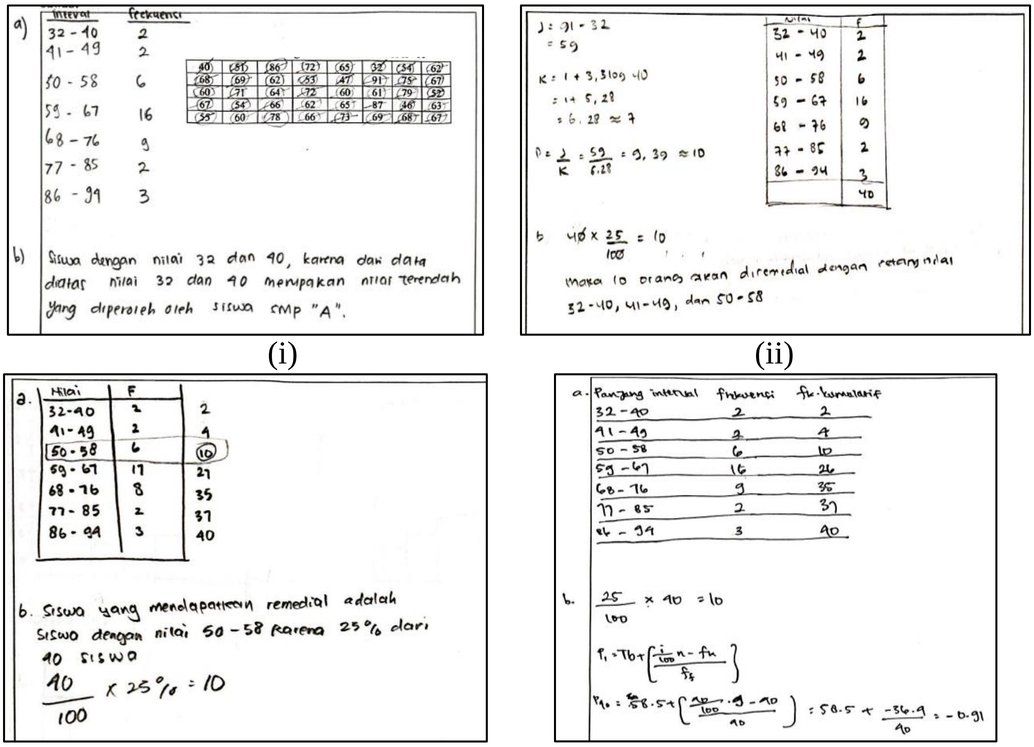
For the second indikator organizing data, students were asked to construct a frequency distribution from a set of mathematics test scores using specified class intervals, and to determine which students should receive remedial instruction based on the lowest 25% of scores. The analysis of students' written responses revealed various difficulties in understanding the concept of data organization, both in constructing fre-

AKSIOMA: Jurnal Program Studi Pendidikan Matematika ISSN 2089-8703 (Print)
Volume 15, No. 2, 2026, 1274-1290 ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

quency distribution tables and in identifying the students eligible for remedial support. Several examples of student answers are presented in Figure 2.

The results of the written test indicate that while some students were able to construct a frequency distribution table, they were not yet able to group data based on positional measures such as deciles or percentiles. As shown in Figure 2 (i), the student relied solely on cumulative frequency, and the answer happened to align with what was required. During the interview, the student admitted not understanding the difference between using cumulative frequency and applying the concept of percentiles. Meanwhile, in Figure 2 (ii), the student applied the steps of Sturges' formula without paying attention to the explicit instruction requiring a specific class interval. Based on the interview, the student explained that the only procedure remembered when constructing a frequency distribution was to use Sturges' method, without considering any additional instructions provided in the question.



(i) (ii) (iii) (iv)

Figure 2 Student Answers to Data Organizing Indicators

Furthermore, in the part of the task requiring students to determine which scores fall into the remedial category, most students relied on cumulative frequency to identify the lowest 25% of scores, rather than using the appropriate positional measure percentiles. This is illustrated in Figure 2 (iii), where the student identified the score range of 50–58 as the remedial threshold, reasoning that this interval contained 10 students. When asked for clarification, the student admitted to not understanding the concept of cumulative frequency and determined the remedial cutoff solely based on the number of students within the interval.

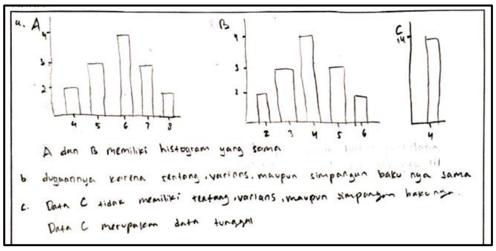
AKSIOMA: Jurnal Program Studi Pendidikan Matematika ISSN 2089-8703 (Print)
Volume 15, No. 2, 2026, 1274-1290 ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

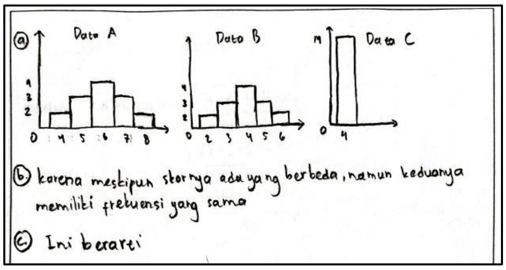
Interestingly, among all participants, only one student attempted to use the percentile concept, although the solution was not completed, as shown in Figure 2 (iv). In the interview, the student explained that they used percentiles because it seemed to be the most reasonable approach, despite being unsure about its accuracy. This finding indicates that most students do not yet possess an adequate understanding of positional measures and their application in organizing data.

The difficulties identified in this indicator can be categorized into several types of learning obstacles. First, the epistemological learning obstacle is evident in students' inability to apply positional measures such as deciles or percentiles. Instead, students tended to rely on cumulative frequency which happened to yield the correct result without understanding the underlying concept. This aligns with Boels et al., (2019) assertion that understanding statistical concepts depends not only on computational procedures but also on grasping the meaning and appropriate use of those concepts within context.

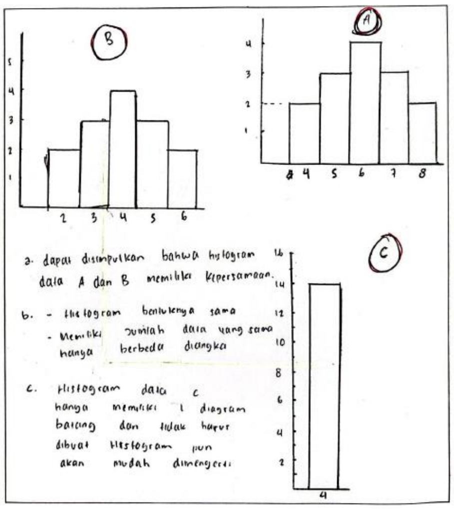
Second, the ontogenic learning obstacle is reflected in students' tendency to rely on previously taught methods such as Sturges' formula without attending to the specific instructions given in the problem. Students' limited readiness to think flexibly and to adapt their strategies to the problem context indicates constraints in their cognitive development and learning experiences. They appear to lack sufficient exposure to applying concepts such as deciles or percentiles in authentic situations, which leads them to avoid using these concepts altogether.



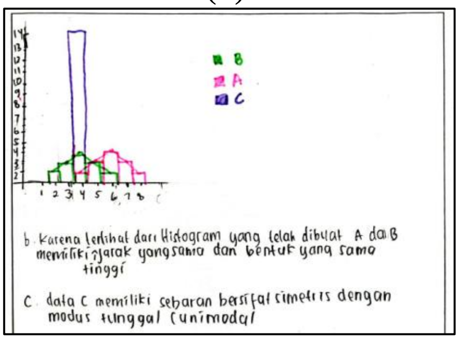
(i)



(ii)



(iii)



(iv)

Figure 3 Student Answers to Indicators Representing Data

AKSIOMA: Jurnal Program Studi Pendidikan Matematika ISSN 2089-8703 (Print)
Volume 15, No. 2, 2026, 1274-1290 ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

Third, a didactical learning obstacle emerges from the instructional approach, learning materials, and classroom interactions. Interview data reveal that during the course, instructional emphasis was placed predominantly on procedural skills for constructing frequency distributions using Sturges' method, while positional measures such as deciles and percentiles were rarely, if ever, integrated into data-organization tasks. Moreover, students did not receive adequate practice connecting positional measures with frequency distributions, nor did they receive targeted feedback to help them correct their misconceptions. Consistent with this, Gravemeijer (1994) argue that the gap between technical procedures and conceptual understanding often constitutes a major source of difficulty in learning statistics.

These findings indicate that students' ability to organize data still requires substantial improvement, particularly in understanding and applying positional measures. An integrative, context-based didactical design is needed to support students in not only mastering computational procedures but also grasping the underlying conceptual meanings and applying these concepts in real world situations.

For the indicator on data representation, students were asked to construct histograms for three data groups (A, B, and C), then explain the similarities in distribution between data A and B, as well as interpret the meaning of data group C, which has a variance and standard deviation of zero. Several examples of student responses are presented in Figure 3. The results show that most students were able to complete part (a) satisfactorily. They successfully constructed histograms for each dataset, indicating that procedurally they understood how to represent data using bar graphs. Interview findings reinforced this result, as students stated that they did not experience significant difficulties in completing part (a).

However, difficulties began to emerge in part (b). Students appeared confused when asked to explain why data A and B shared the same distribution despite having different individual scores. Most students relied solely on the visual appearance of the histograms without understanding that two datasets can exhibit identical relative frequency distributions even when their values differ. During interviews, students reported that, aside from forgetting the concept, they were unsure how to respond; consequently, many answers were based merely on subjective impressions of the histograms rather than deeper analytical reasoning.

A similar pattern occurred in part (c). Students struggled to explain the meaning of data group C, which had a variance and standard deviation of zero. They were unable to recognize that these values indicate complete homogeneity in the dataset, where all observations are identical and therefore exhibit no variability. When asked for clarification during the interviews, students stated that the question was confusing and that they did not understand the implication of having a variance and standard deviation of zero.

The difficulties identified in this indicator can be categorized into several types of learning obstacles. First, an epistemological learning obstacle is evident from students' misconceptions regarding the concept of data distribution. Students did not understand that identical relative frequency distributions can occur in datasets with different individual values. In addition, they had

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

not yet mastered the conceptual meaning of a variance and standard deviation of zero.

Second, an ontogenic learning obstacle emerges from students' prior learning experiences, which tended to emphasize technical skills in constructing visual representations rather than providing sufficient focus on analyzing the properties of data distributions. As a result, students were not accustomed to engaging in deeper interpretations of the data they represented.

Third, a didactical learning obstacle is assumed to stem from instructional practices that place greater emphasis on the procedural steps of constructing histograms than on developing students' interpretive abilities. The practice problems provided also appeared insufficiently varied and did not adequately encourage students to understand concepts related to data dispersion both conceptually and applicatively.

For the indicator of analyzing and interpreting data, students were asked to determine the standard deviation for two datasets (Mathematics and English) and to identify in which subject a student named Budi demonstrated better performance, based on a comparison between his individual score and the mean score of each subject. Several examples of students' responses are presented in Figure 4. The findings indicate that most students experienced difficulties in solving this problem. As shown in Figures 4 (i) and 4 (ii), students performed calculations using unclear and incorrect formulas, suggesting that they did not understand the proper procedure for computing the standard deviation. Interview results reinforced this conclusion; students admitted that they were confused and simply wrote down calculations based on the numbers provided in the problem, without understanding the underlying concepts or correct steps.

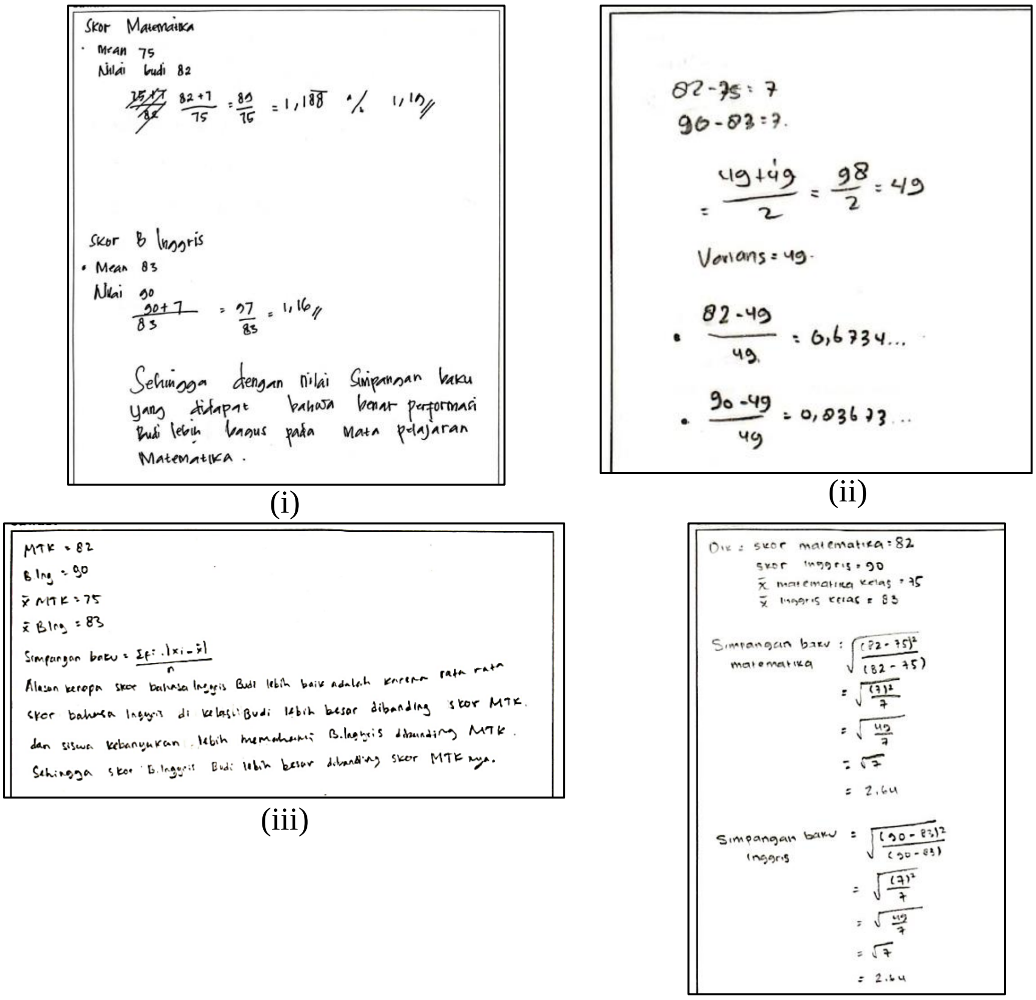
In Figure 4 (iii), the student only compared the mean scores of the two groups without considering the standard deviation, demonstrating a lack of understanding regarding the importance of measures of dispersion when evaluating relative performance. The response in Figure 4 (iv) shows a student attempting to compute the standard deviation for both groups, but stopping at the calculation and failing to connect it to Budi's relative performance. This indicates that although the student recognized the relevance of standard deviation in data analysis, they had not yet developed the ability to relate the computed values to the interpretation required by the task.

The difficulties encountered by the students in solving this problem can be categorized into several types of learning obstacles. First, an epistemological learning obstacle is evident in students' lack of understanding of fundamental statistical concepts, particularly in selecting and applying the appropriate formula for standard deviation. Students also did not grasp the importance of comparing individual scores relative to the data's dispersion when determining performance. Second, an ontogenic learning obstacle is reflected in students' limited familiarity with tasks requiring them to relate standard deviation to concepts such as Z-scores or relative performance comparisons. Their prior learning experiences, which tended to emphasize mean calculations without incorporating measures of dispersion, are likely a primary source of this difficulty. Third, a didactical learning obstacle arises from instructional approaches that appear to rely more

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

heavily on simple problems that use the mean as the sole indicator of performance. Consequently, students are unaccustomed to more complex tasks that require integrating measures of dispersion into performance analysis.



(i) (ii) (iii) (iv)

Figure 4. Student Answers to Indicators Representing Data

Based on the results of the students' test performance across the four indicators, several learning obstacles were identified, particularly in their understanding of fundamental statistical concepts. The obstacles that emerged were predominantly epistemological learning obstacles, characterized by misconceptions regarding measures of position, data variability, and relative performance. In addition, ontogenic learning obstacles were observed in students' tendency to rely on previously learned procedures without adjusting them to the context of the problem. Meanwhile, the didactical learning obstacles appear to stem from instructional approaches that place greater emphasis on procedural aspects rather than on conceptual and analytical understanding. This preliminary conclusion serves as the basis for further exploration through interviews with the course instructor, as well as a document analysis of the teaching materials and students' notes.

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

Interview results with the course instructor indicate that students have primarily been exposed to procedural approaches when solving statistical problems, particularly through the application of formulas and numerical computations. This finding is consistent with previous studies showing that statistics instruction often emphasizes computational procedures rather than conceptual understanding and statistical reasoning (Garfield et al., 2008). Recent studies similarly report that students frequently develop procedural proficiency without adequately understanding the underlying statistical concepts, which limits the development of statistical thinking and reasoning skills (Azmay et al., 2023; Hasim & Rosli, 2024). Consequently, students tend to focus on obtaining numerical answers while experiencing difficulties in interpreting and connecting statistical concepts across different contexts. The limited emphasis on conceptual understanding and interpretation therefore appears to be a major factor contributing to the emergence of learning obstacles.

In addition, measures of position such as deciles and percentiles are often introduced only briefly and with limited opportunities for contextual application. Del-Callejo-Canal et al., (2020) emphasized that understanding measures of position and dispersion is essential for developing holistic statistical thinking. This view is supported by recent research indicating that meaningful statistical learning requires students to engage with authentic data, contextual reasoning, and interpretation rather than relying solely on computational procedures (Azmay et al., 2023; Schreiter et al., 2024). The lack of contextualized learning experiences may therefore contribute to students' difficulties in applying positional measures appropriately.

The document analysis revealed that instructional materials were dominated by procedural explanations and formula based exercises. Similar findings have been reported in recent studies showing that statistics learning resources often provide insufficient support for conceptual understanding, data interpretation, and statistical literacy (Engel, 2017; Schreiter et al., 2024). Furthermore, students' notes tended to focus primarily on formulas and calculation procedures rather than conceptual explanations, indicating that learning activities may not adequately promote reflection on the meaning of statistical concepts. This pattern aligns with Fielding et al. (2025), who noted that data-handling instruction has often been limited to calculations, graphing, and interpretation, whereas meaningful engagement with data requires students to consider data sources, contexts, purposes, and implications.

A summary of the findings from the lecturer interviews and the document analysis is presented in Table 1.

Table 1. Summary of Lecturer Interviews and Documentation Analysis

No	Data Source	Key Findings	Information
1	An Interview with the Lecturer	Students tend to only follow the calculation procedure without understanding the meaning of the calculation results	Support findings on indicators, describe data and represent data

AKSIOMA: Jurnal Program Studi Pendidikan Matematika
Volume 15, No. 2, 2026, 1274-1290
ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

No	Data Source	Key Findings	Information
2	An Interview with the Lecturer	The material on location size (decils/percentiles) is only presented concisely, the questions are rarely contextual	Regarding student barriers in indicators organizing data
3	Teaching Material Documentation Study	Teaching materials contain formulaic procedures without emphasis on data interpretation	Didactical obstacle dalam pembelajaran statistika
4	Student Record Documentation Study	Student notes focus on calculation formulas, the concept of the meaning of variances and standard deviations are rarely recorded	In accordance with the results of tests and student interviews in the indicator of analyzing and interpreting data

The findings of this study demonstrate that prospective mathematics teachers experience learning obstacles across all indicators of statistical thinking, including describing data, organizing data, representing data, and analyzing and interpreting data. Among the three categories of learning obstacles, epistemological obstacles were found to be the most dominant. Students frequently demonstrated misconceptions regarding measures of central tendency, measures of position, data variability, and relative performance. These findings suggest that students' difficulties are not merely procedural errors but reflect deeper conceptual misunderstandings of fundamental statistical ideas.

The predominance of epistemological obstacles may be explained by students' fragmented understanding of statistical concepts. Although many students were able to perform calculations, they struggled to connect numerical procedures with graphical representations and contextual interpretations. For example, students were unable to relate the mean, median, and mode to distribution skewness, distinguish between the position and value of the median, or interpret the meaning of a standard deviation equal to zero. This indicates that statistical concepts were learned as isolated procedures rather than as interconnected concepts within a coherent statistical framework.

The findings are consistent with previous studies reporting students' difficulties in understanding statistics beyond procedural computation. Silva et al., (2025), based on a survey of statistics instructors, reported that college students often experience difficulties with conceptual understanding, connecting interrelated statistical ideas, and applying statistical knowledge to problem-solving situations. These findings suggest that students' challenges in statistics extend beyond carrying out procedures and involve the abstraction and interpretation of statistical concepts. Similarly, Schreiter & Vogel, (2025) found that students often relied on local and computational approaches when comparing data distributions, rather than interpreting distributions holistically through their center, spread, and shape.

Several factors appear to contribute to the emergence of these learning obstacles. Interview results with the course instructor and document analysis revealed that instruction tended to emphasize formula application and

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

numerical computation rather than conceptual understanding and data interpretation. Learning materials focused primarily on procedural exercises, while opportunities for contextual reasoning, visual interpretation, and statistical communication were limited. Consequently, students became accustomed to applying memorized procedures without fully understanding the underlying statistical concepts.

A notable contribution of this study is its integration of the statistical thinking framework with the learning obstacle perspective. While previous studies have generally examined students' difficulties in learning statistics, this study provides a more comprehensive explanation by classifying those difficulties into epistemological, ontogenic, and didactical obstacles. This classification offers deeper insight into the sources of students' difficulties and provides a stronger theoretical basis for instructional improvement.

The findings also have important implications for statistics education in teacher preparation programs. Learning activities should place greater emphasis on conceptual understanding, interpretation of statistical representations, and contextual problem solving rather than merely procedural calculations. Furthermore, the identified learning obstacles can serve as an empirical foundation for developing hypothetical learning trajectories and didactical designs aimed at fostering students' statistical thinking in descriptive statistics courses.

CONCLUSIONS

This study aimed to identify the learning obstacles experienced by prospective mathematics teachers in descriptive statistics courses. The findings revealed that students encountered learning obstacles across all indicators of statistical thinking, namely describing data, organizing data, representing data, and analyzing and interpreting data. The identified obstacles consisted of epistemological, ontogenic, and didactical learning obstacles, with epistemological obstacles being the most dominant. These obstacles were reflected in students' misconceptions regarding measures of central tendency, measures of position, data variability, and relative performance.

The study contributes to the literature by providing a systematic identification of learning obstacles based on statistical thinking indicators and by classifying these obstacles through the framework of epistemological, ontogenic, and didactical obstacles. The findings provide an empirical foundation for the development of didactical designs and hypothetical learning trajectories aimed at improving students' statistical thinking in descriptive statistics courses.

Future studies are recommended to develop and implement didactical designs based on the identified learning obstacles and to investigate their effectiveness in enhancing students' statistical thinking abilities.

REFERENCES

Artigue, M. (2009). *Didactical design in mathematics education*. Brill Sense. https://doi.org/https://doi.org/10.1163/9789087907839_003

Azis, A., & Dahlan, J. A. (2024). Statistical Literacy Level of Mathematics Education Students : Challenges and Recommendations for Competency Improvement. *Alifmatika: Jurnal Pendidikan Dan Pembelajaran Matematika*, *6*(2), 263–278.

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

https://doi.org/10.35316/alifmatika.2024.v6i2.263-278

Azmay, N. A. M. N., Rosli, R., Maat, S. M., & Mahmud, M. S. (2023). Educational Research Trends on Statistical Reasoning and Statistical Thinking : A Systematic Literature Review. *International Journal Of Academic Research In Progressive Education And Development*, *12*(2), 543–556. https://doi.org/10.6007/IJARPED/v12-i2/16921

Ben-Zvi, D., & Makar, K. (2016). International Perspectives on the Teaching and Learning of Statistics BT - The Teaching and Learning of Statistics: International Perspectives. In D. Ben-Zvi & K. Makar (Eds.), *The Teaching and Learning of Statistics* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-23470-0_1

Boels, L., Bakker, A., Dooren, W. Van, & Drijvers, P. (2019). Conceptual difficulties when interpreting histograms : A review. *Educational Research Review*, *28*(October 2018), 100291. https://doi.org/10.1016/j.edurev.2019.100291

Brousseau, G. (2002). *Theory Of Didactical Situations In Mathematics Didactique Des Mathématiques, 1970–1990*. Kluwer Academic Publishers.

Chaphalkar, R., & Wu, K. (2020). Students ' Reasoning about Variability in Graphs during an Introductory Statistics Course. *International Electronic Journal of Mathematics Education*, *15*(2), 1–22. https://doi.org/https://doi.org/10.29333/iejme/7602

Cipta, E. S., Suryadi, D., Herman, T., & Jupri, A. (2024). The utilization of technology for optimizing statistical thinking: Global trends and directions for future research. *AIP Conference Proceedings*, *3220*(1). https://doi.org/https://doi.org/10.1063/5.0235077

Cipta, E. S., Suryadi, D., Herman, T., & Jupri, A. (2026). *The Effect Of Digital Didactic Intervention On Measures Of Central Tendency Towards Students ' Statistical Thinking*. *9*(1), 164–176. https://doi.org/https://dx.doi.org/10.22460/jiml.v9i1.p29891

Cipta, E. S., Suryadi, D., Herman, T., Jupri, A., Konidah, E., Chusaery, H. A. L., Haryadi, H., & Sarinah, S. (2024). Enhancing Students' Understanding of the Mean Concept Through Problem-Based Learning Assisted By Geogebra: a Quasi-Experimental Study. *Journal of Engineering Science and Technology*, *19*(5), 18–26. https://jestec.taylors.edu.my/Special Issue ISCoE 2023_3/ISCoE 2023 3_03.pdf

Del-Callejo-Canal, D., Canal-Martínez, M., & Hákim-Krayem, M. R. (2020). Statistical thinking development in superior level students through one educative experience. *Educacion Matematica*, *6*(2), 194–216. https://doi.org/10.24844/EM3202.08

Engel, J. (2017). Statistical Literacy For Active Citizenship : A Call For Data Science Education. *Statistics Education Research Journal*, *16*(1), 44–49. https://doi.org/http://iase-web.org/Publications.php?p=SER

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

J

Fielding, J., Makar, K., & Ben, D. (2025). Developing students ' reasoning with data and data - ing. *ZDM – Mathematics Education*, *57*(1), 1–18. https://doi.org/10.1007/s11858-025-01671-6

Friedrich, A., Schreiter, S., Vogel, M., Genschow, S. B., Brünken, R., Kuhn, J., Lehmann, J., & Malone, S. (2024). What shapes statistical and data literacy research in K - 12 STEM education ? A systematic review of metrics and instructional strategies. *International Journal of STEM Education*. https://doi.org/10.1186/s40594-024-00517-z

Frischemeier, D., & Schnell, S. (2023). Statistical investigations in primary school – the role of contextual expectations for data analysis. *Mathematics Education Research Journal*, *35*(1), 217–242. https://doi.org/10.1007/s13394-021-00396-5

Garfield, J., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice. In *Springer*. https://doi.org/10.1007/978-1-4020-8383-9

Gravemeijer, K. P. E. (1994). *Developing Realistic Mathematics Education*.

Hasim, S. M., & Rosli, R. (2024). A Systematic Review on Teaching Strategies for Fostering Students ' Statistical Thinking. *International Journal of Learning, Teaching and Educational Research*, *23*(1), 136–158. https://doi.org/https://doi.org/10.26803/ijlter.23.1.8

Hidayat, W., Widodo, S. A., & Syahrizal, T. (2023). The statistical thinking skill and adversity quotient of English pre-service teacher. *International Journal of Evaluation and Research in Education*, *12*(1), 421–432. https://doi.org/10.11591/ijere.v12i1.24302

Moleong. (2007). *Metodologi penelitian kualitatif*. PT Remaja Rosdakarya.

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2024). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, *22*(October), 1–18. https://doi.org/10.1177/16094069231205789

Oslington, G., Mulligan, J., & Bergen, P. Van. (2024). Shifts in Students' Predictive Reasoning from Data Tables in Years 3 and 4. *Mathematics Education Research Journal*, *36*(3), 547–576. https://doi.org/10.1007/s13394-023-00460-2

Patton, M. Q. (2022). *Qualitative research and evaluation methods (3rd ed.)*. SAGE Publications.

Prodromou, T., & Dunne, T. (2017). Statistical literacy in data revolution era: Building blocks and instructional dilemmas. *Statistics Education Research Journal*, *16*(1), 38–43. https://doi.org/10.52041/serj.v16i1.212

Rohimah, S. M., Putri, S. A., & Helsa, Y. (2026). Conceptual and procedural errors of pre-service elementary teachers in solving

ISSN 2089-8703 (Print)
ISSN 2442-5419 (Online)

DOI: https://doi.org/10.24127/ajpm.v15i2.14702

educational research statistics problems. *Jurnal Elemen*, *12*(1), 49–65. https://doi.org/https:/doi.org/10.29408/jel.v12i1.32327

Şahal, M., & Koklu, O. (2026). Examining Pre - service Mathematics Teachers ' Statistical Investigations : Challenges and Instructional Support. In *International Journal of Science and Mathematics Education* (Vol. 24, Issue 19). Springer Nature Singapore. https://doi.org/10.1007/s10763-025-10634-y

Sanchez, J. M. (2023). The Need to Reinforce the Teaching of Basic Descriptive Statistics Required in Reporting Quantitative Laboratory Results: Diagnose of Common Students' Misconceptions. *Journal of Chemical Education*, *100*(7), 2713–2718. https://doi.org/10.1021/acs.jchemed.3c00394

Schreiter, S., Friedrich, A., Fuhr, H., Malone, S., Brünken, R., Kuhn, J., & Vogel, M. (2024). Teaching for statistical and data literacy in K - 12 STEM education : a systematic review on teacher variables , teacher education , and impacts on classroom practice. *ZDM – Mathematics Education*, *56*(1), 31–45. https://doi.org/10.1007/s11858-023-01531-1

Schreiter, S., & Vogel, M. (2025). Students ' local vs . global views of data distributions : a cross - grade - level analysis using eye - tracking. *Educational Studies in Mathematics*, *119*, 1–22. https://doi.org/10.1007/s10649-024-10352-2

Setiawan, E. P., & Sukoco, H. (2021). Exploring First Year University Students' Statistical Literacy: A Case on Describing and Visualizing Data. *Journal on Mathematics Education*, *12*(3), 427–448. https://doi.org/http://doi.org/10.22342/jme.12.3.13202.427-448

Silva, P. N., Sarnecka, B. W., & Sarnecka, B. W. (2025). What Do Your Students Struggle with ? A Survey of Statistics Instructors What Do Your Students Struggle with ? A Survey of Statistics Instructors ABSTRACT. *Journal of Statistics and Data Science Education*, 1–12. https://doi.org/10.1080/26939169.2025.2455560

Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D*. Alfabeta.

Suryadi, D. (2019). *Landasan Filosofis Penelitian Desain Didaktis (DDR)*. Pusat Pengembangan DDR Indonesia.