Volume 8, Number 2, October 2026
Jurnal Didactical Mathematics
DOI: https://doi.org/10.31949/dm.v8i2.19208
p-ISSN: 2622-7525, e-ISSN: 2654-9417

Stage-Specific Patterns in Students' Statistical Problem Solving: A Qualitative Analysis Through Pólya's Framework

Laura Fatikhatus Salim[1]*, Muthiah Fildzah Noverli[2], Hendra Kartika[3]
[1,2,3] Department of Mathematics Education, Universitas Singaperbangsa Karawang

*Corresponding Author: 2210631050123@student.unsika.ac.id

Article Info

Article History
Submitted: 07-07-2026
Revised: 01-09-2026
Accepted: 03-09-2026
Published: 04-09-2026

Keywords:

Mathematical problem solving;
Statistical problem solving;
Pólya's problem-solving stages;
Process-oriented assessment;
Grade VIII students

Abstract

This study examined Grade VIII students' mathematical problem-solving performance in statistical tasks through Pólya's four problem-solving stages: understanding the problem, devising a plan, carrying out the plan, and looking back. A qualitative descriptive design was employed to examine students' written responses across these stages. The participants were 30 eighth-grade students from MTsN 4 Bogor who completed an open-ended mathematical problem-solving test. Based on their overall scores, six students representing high-, moderate-, and low-achievement categories were purposively selected for qualitative analysis. The analysis focused on the evidence of students' problem interpretation, planning, procedural execution, and post-solution evaluation in their written work. The findings showed that students' performance was uneven across the four stages. Students differed in how explicitly they identified relevant information and documented their solution plans, while procedural execution varied in accuracy and consistency. The clearest difficulty occurred in the *looking-back* stage, as the responses examined did not provide explicit written evidence of verification or evaluation of the obtained results. Stage-specific performance also did not consistently correspond to overall achievement classification: students with lower overall scores demonstrated competence in particular stages, while a student with the highest overall score did not explicitly demonstrate verification in the response examined. These findings indicate that overall achievement scores do not fully represent the characteristics of students' problem-solving processes. Examining written work across Pólya's stages can therefore provide a more differentiated account of students' statistical problem solving, particularly by making stage-specific strengths and difficulties visible.

INTRODUCTION

Mathematical problem solving is a central component of mathematics education because it requires students to coordinate mathematical knowledge, representations, strategies, and reasoning rather than merely produce correct answers. Contemporary research has increasingly examined problem solving as a process, focusing on how students interpret problems, construct solutions, and evaluate their reasoning rather than on final answers (Liljedahl & Cai, 2021). In this process, students may coordinate different forms of representation, including symbolic and verbal representations, as they interpret information and develop mathematical solutions (Anwar & Rahmawati, 2017). Mathematical problem-solving performance is therefore multidimensional, involving not only procedural knowledge but also cognitive and metacognitive processes that influence how students approach and regulate their solutions (Özcan & Eren Gümüş, 2019; Güner

& Erbay, 2021). This process-oriented perspective matters for didactical analysis because an overall measure of achievement may not fully capture how students engage with specific components of mathematical problem solving.

Pólya's four-stage framework provides an established way to examine these processes by understanding the problem, devising a plan, carrying out the plan, and looking back. Rather than treating problem solving as a single outcome, the framework enables the analysis of how students move through interconnected stages of mathematical activity (Kania & Ratnawulan, 2022). Previous studies have shown that students' performance can differ across these stages, with difficulties emerging in interpreting problem information, selecting appropriate strategies, and carrying out or completing solutions (Luthfiyah et al., 2022). Research has also shown that students' representations may develop across the stages of Pólya's framework, further indicating that problem solving involves a sequence of related reasoning and representational processes (Anwar & Rahmawati, 2017). Thus, Pólya's framework is useful not only for classifying students according to overall problem-solving ability but also for examining where particular difficulties occur within the solving process. This perspective aligns with evidence that mathematical problem-solving performance cannot be adequately understood through a single dimension of achievement because different cognitive and metacognitive processes may operate during problem solving (Özcan & Eren Gümüş, 2019; Güner & Erbay, 2021).

This process-oriented view is particularly relevant to statistical problem solving, where students must interpret information and representations, organize and analyze data, apply appropriate procedures, and make sense of the resulting information. Statistical problem solving therefore involves more than procedural calculation and requires students to coordinate statistical information with mathematical reasoning (Rosidah et al., 2018; Hidayanto & Rahmatina, 2020). Recent research has examined statistical problem solving through stage-based approaches, including using Pólya's framework to investigate students' statistical problem-solving skills (Makwakwa et al., 2024). Research involving middle school students has also indicated that components of the statistical problem-solving process may receive insufficient attention when emphasis is placed primarily on the final solution (Hedges et al., 2022). Within the Indonesian context, students' problem solving in statistical settings has been examined through Pólya's stages, including research on statistical problems in problem-based learning (Aini et al., 2023) and problem solving involving data presentation (Meisy et al., 2022). These studies establish statistical problem solving as a context in which students' processes of interpretation, planning, execution, and evaluation can be examined systematically.

Although this literature provides substantial evidence concerning students' performance and difficulties across Pólya's stages, an important process-level issue remains. Existing studies have often described students' problem-solving abilities according to the four stages or examined differences by mathematical topic and achievement level (Aini et al., 2023; Luthfiyah et al., 2022). Such approaches provide useful information about students' strengths and difficulties, but they do not necessarily establish whether an overall achievement classification consistently represents performance across each individual stage. This distinction matters because problem-solving processes may develop unevenly: a student may successfully execute a mathematical procedure while showing limited evidence of evaluating the solution, or may demonstrate adequate understanding without developing a coherent plan. Consequently, examining students only through an overall classification of high, moderate, or low achievement may obscure meaningful differences in how particular problem-solving processes are enacted. A process-level examination is therefore needed to determine how performance at individual stages corresponds to students' overall achievement in statistical problem solving.

The *looking-back* stage is especially important in examining this issue because it involves checking the solution, reconsidering the reasoning, and evaluating whether the result is appropriate for the problem. Research on students' use of look-back strategies shows that looking back can support reconsidering solutions and exploring alternative solution methods (Lee, 2016). Similarly, Koichu et al. (2021) examined looking back as a mathematical practice and highlighted the importance of supporting students in engaging in this form of reflection. The relevance of looking back is also connected with metacognitive processes in mathematical problem solving, particularly monitoring and evaluating performance after a solution has been produced (García et al., 2016; Güner & Erbay, 2021). These perspectives indicate that completing a procedure successfully does not necessarily mean that students have explicitly examined or verified the resulting solution. However, less is known about how students' engagement in looking back relates to their performance in the preceding Pólya stages and to their overall achievement when solving statistical problems. Examining this relationship can therefore provide a more differentiated account of students' problem-solving processes than an overall achievement score alone.

Against this background, the present study examines eighth-grade students' mathematical problem-solving performance in statistical tasks through Pólya's four stages: understanding the problem, devising a plan, carrying out the plan, and looking back. The analysis focuses on how students' performance varies across these stages in relation to their overall achievement levels. Rather than using the Pólya stages only to assign an overall problem-solving category, the study examines students' responses at each stage to identify patterns in the processes underlying their statistical problem solving. Particular attention is given to the *looking-back* stage and its relationship with performance in the preceding stages, providing a more detailed examination of how students' problem-solving performance manifests across different levels of overall achievement.

METHODS

Research Design

This study employed a qualitative descriptive design to examine eighth-grade students' mathematical problem-solving performance when solving statistical tasks through Pólya's four stages: understanding the problem, devising a plan, carrying out the plan, and looking back. A qualitative descriptive design was selected because the study aimed to provide a systematic and empirically grounded description of students' problem-solving processes without testing causal relationships or evaluating the effectiveness of an instructional intervention. Qualitative descriptive research is particularly appropriate when researchers seek to present a phenomenon in a relatively direct and accessible form while maintaining a close connection to the empirical data (Villamin et al., 2025). The design therefore allowed the analysis to remain closely grounded in students' written solutions and the observable characteristics of their responses at each problem-solving stage. This orientation is consistent with mathematics education research that treats mathematical activity as situated in students' engagement with tasks and the ways in which they coordinate mathematical resources during problem solving (Abrahamson & Sánchez-García, 2016). Accordingly, the study focused on describing stage-specific patterns of problem-solving performance rather than inferring unobserved cognitive processes.

Research Participants

The study was conducted at MTsN 4 Bogor, Indonesia, and involved 30 eighth-grade students from one intact classroom. The class was purposively selected because it represented the target population and was accessible for study implementation. After administering the mathematical problem-solving test, students were classified into three achievement categories—high, moderate, and low—based on their overall problem-solving scores. Two students from each category were then purposively selected as focal participants, resulting in six students for the qualitative analysis. The selection was intended to enable comparison of students' problem-solving processes across different overall achievement levels rather than to produce statistical generalization to the wider student population.

Research Instrument

The primary research instrument was an open-ended mathematical problem-solving test focusing on statistical concepts. The tasks were designed to elicit written evidence of students' problem-solving processes based on Pólya's four stages: (1) understanding the problem, (2) devising a plan, (3) carrying out the plan, and (4) looking back. Students were instructed to identify the information provided in the problem, formulate an appropriate strategy, present the procedures used to obtain a solution, and evaluate the reasonableness of their final answers. The use of open-ended tasks was intended to make students' mathematical problem-solving processes visible through their written responses rather than relying solely on the correctness of the final answer. This approach is consistent with the view that mathematical task design can provide opportunities for students to demonstrate and communicate their mathematical reasoning and problem-solving activity (Chan & Clarke, 2017). Pólya's framework was subsequently used as the analytical basis for examining the extent to which students demonstrated each stage of the problem-solving process in their written work. Before data collection, experts reviewed the instrument to assess the relevance and clarity of the tasks and their alignment with the four Pólya problem-solving indicators. The expert review ensured that the tasks were appropriate for eliciting evidence of the intended problem-solving processes. The students' written responses were then collected and retained as the primary empirical data for analysis according to the four stages of Pólya's problem-solving framework.

Data Collection Procedures

Data collection was conducted sequentially. First, permission to conduct the study was obtained from the school administration, followed by coordination with the mathematics teacher concerning the research schedule and classroom implementation. The mathematical problem-solving test was subsequently administered to the 30 students during regular classroom instruction. The test consisted of open-ended statistical tasks structured to elicit evidence corresponding to the four Pólya stages. Students were instructed to write their solution procedures explicitly, including their interpretation of the problem, planned strategy, mathematical procedures, and evaluation of the obtained solution. The written responses were collected after completion of the test and subsequently used for scoring and qualitative analysis.

Scoring and Classification of Mathematical Problem-Solving Performance

Students' mathematical problem-solving performance was evaluated using an analytical scoring rubric based on Pólya's four stages: (1) understanding the problem, (2) devising a plan, (3) carrying out the plan, and (4) looking back. Each indicator was scored on a three-point scale ranging from 0 to 2. A score of 0 indicated that no appropriate response or relevant evidence was provided, a score of 1 indicated a partially correct, incomplete, or inaccurate response, and a score of 2 indicated an appropriate and complete response. The scoring focused on the evidence of students' problem-solving processes demonstrated in their written responses rather than solely on the correctness of the final answer. The scores across the four indicators were summed to obtain each student's overall problem-solving score and then converted to a 0–100 scale. Based on the predetermined classification criteria, students were categorized as high (75–100), moderate (50–74), or low (0–49). These categories were used to identify focal participants for the subsequent qualitative analysis. Accordingly, the overall score functioned as a basis for participant selection, while the students' responses at each stage of Pólya's problem-solving framework served as the primary evidence for examining differences in their mathematical problem-solving processes.

Table 1. Classification of Students' Mathematical Problem-Solving Performance

Score	Category
75–100	High
50–74	Moderate
0–49	Low

Table 2. Scoring Rubric Based on Pólya's Problem-Solving Stages

No.	Indicator	Performance Description	Score
1	Understanding the Problem	Does not identify the known or required information or provides an irrelevant response.	0
		Identifies some of the known and required information, but incompletely or inaccurately.	1
		Correctly and completely identifies the known and required information.	2
2	Devising a Plan	Does not formulate an appropriate solution strategy.	0
		Formulates a strategy, but it is incomplete or inappropriate.	1
		Formulates an appropriate and complete solution strategy.	2
3	Carrying Out the Plan	Does not attempt the solution or applies an inappropriate procedure.	0
		Applies an appropriate procedure but commits computational or procedural errors.	1
		Correctly implements the procedure and obtains an accurate solution.	2
4	Looking Back	Does not review or evaluate the obtained solution.	0
		Reviews the solution, but the verification is incomplete or inaccurate.	1
		Thoroughly verifies the solution and provides an appropriate conclusion.	2

Qualitative Analysis of Stage-Specific Problem-Solving Performance

Data analysis was conducted in two related phases. First, the responses of all 30 students were scored using the Pólya-based rubric and converted to the overall achievement scale described above. Second, the written responses of the six focal participants were examined qualitatively across the four Pólya stages. For each participant, the analysis identified evidence of understanding the problem, devising a plan, carrying out the plan, and looking back. Attention was given to the completeness, appropriateness, and accuracy of the evidence demonstrated at each stage. The analysis then compared patterns within and across the high-, moderate-, and low-achievement groups. Particular attention was given to whether students' performance at individual stages was consistent with their overall achievement classification. This comparison helped identify similarities, differences, and instances of uneven performance across the problem-solving process. The analysis remained grounded in observable features of students' written responses; therefore, interpretations concerning reflection, monitoring, or verification were based on explicit evidence in the responses rather than on assumptions about students' internal cognitive states. This approach is consistent with qualitative descriptive analysis, which emphasizes a close connection between empirical data and the resulting description of the phenomenon (Villamin et al., 2025).

Ethical Considerations

The researchers followed ethical procedures throughout the study. The researchers obtained permission to conduct the research from the administration of MTsN 4 Bogor, and participation was voluntary. Students were informed about the purpose and procedures of the study and that their responses would be used solely for research purposes. Participants' identities were protected through anonymized identification codes, and no personally identifiable information was disclosed in reporting the findings.

RESULT

Overall Achievement Profile and Selection of Focal Participants

The mathematical problem-solving test was administered to 30 eighth-grade students. Students' overall performance was first evaluated using the Pólya-based scoring rubric described in the Methods. The resulting scores were converted to a 0–100 scale and used to classify students into high, moderate, and low achievement categories. Six students were subsequently selected as focal participants for detailed qualitative analysis of their written responses across the four Pólya stages.

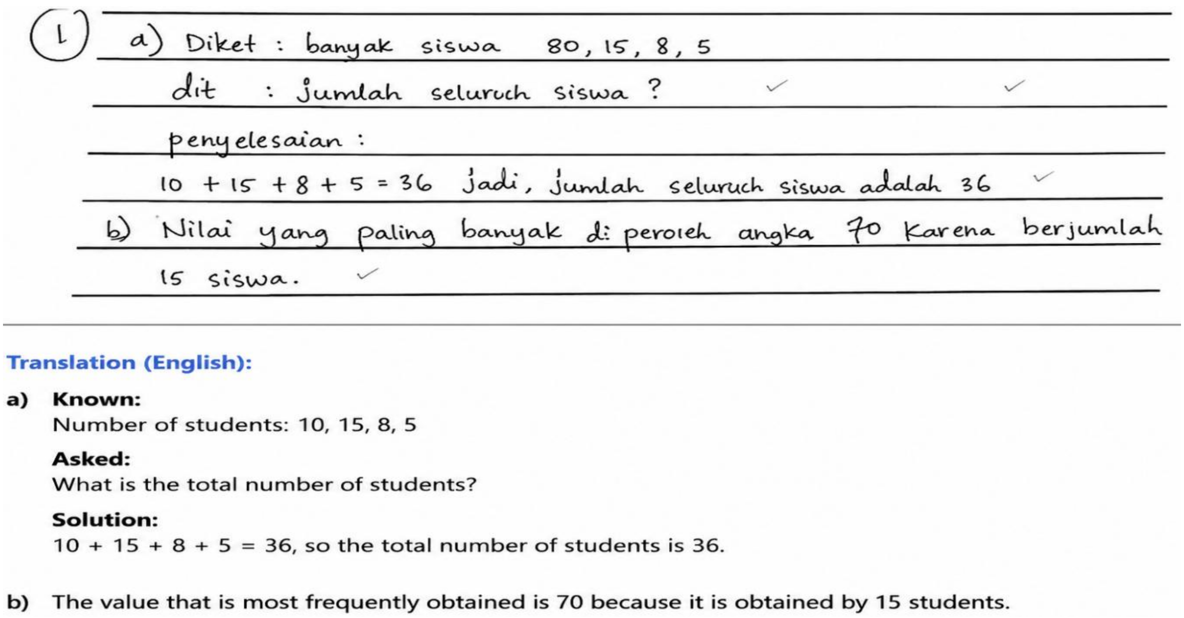
Table 3. Overall Mathematical Problem-Solving Scores of Focal Participants

No.	Subject Code	Total Score	Category
1	S22	100	High
2	S26	87.5	High
3	S17	62.5	Moderate
4	S32	75	**High**
5	S10	50	**Moderate**
6	S20	37.5	Low

The six focal participants provided variation in overall mathematical problem-solving performance, with scores ranging from 37.5 to 100. Their written responses were subsequently examined across the four Pólya stages: understanding the problem, devising a plan, carrying out the plan, and looking back. The stage-level analysis focused on the evidence explicitly demonstrated in the students' written work rather than on their final answers alone.

Understanding the Problem

The *understanding the problem* stage concerns students' identification of relevant information presented in a task and determination of the information required before proceeding to a solution. Analysis of the three written responses revealed differences in how these elements were explicitly represented, particularly in the organization of the given information, identification of the required quantities, and transition from problem information to the resulting statistical answer.



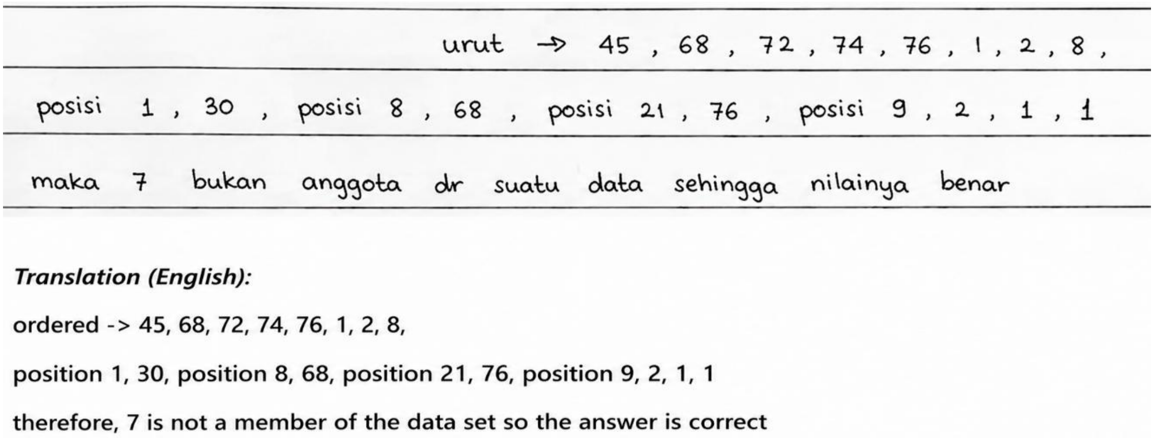
1) a) Diket : banyak siswa 80, 15, 8, 5
dit : Jumlah seluruh siswa ?
penyelesaian :
10 + 15 + 8 + 5 = 36 Jadi, Jumlah seluruh siswa adalah 36
b) Nilai yang paling banyak di peroleh angka 70 Karena berjumlah 15 siswa.

Translation (English):

a) **Known:**
Number of students: 10, 15, 8, 5
Asked:
What is the total number of students?
Solution:
10 + 15 + 8 + 5 = 36, so the total number of students is 36.

b) The value that is most frequently obtained is 70 because it is obtained by 15 students.

Figure 1. Response of S26 to Problem 1

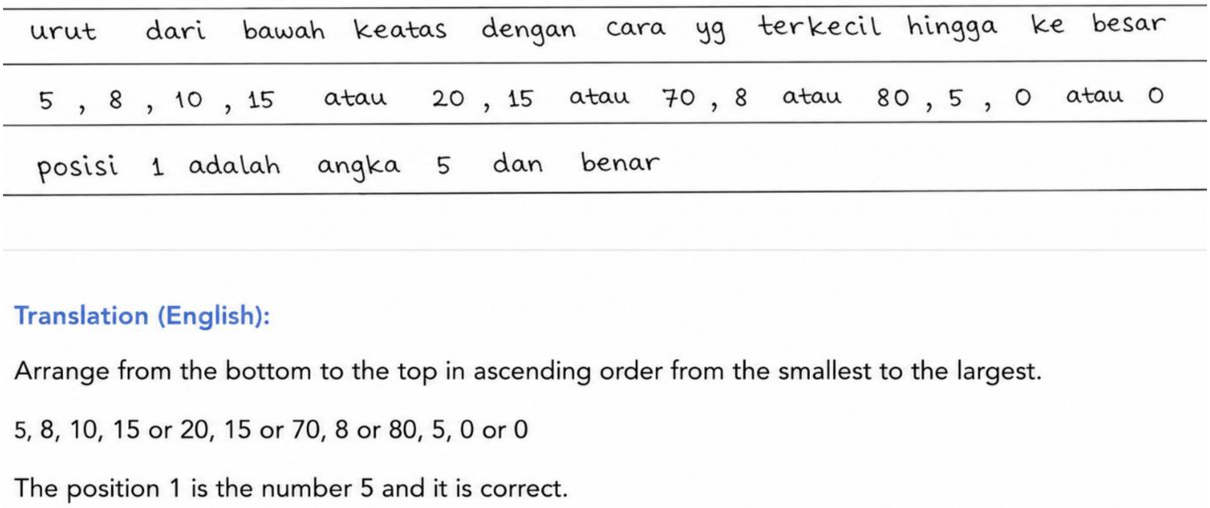


urut → 45 , 68 , 72 , 74 , 76 , 1 , 2 , 8 ,
posisi 1 , 30 , posisi 8 , 68 , posisi 21 , 76 , posisi 9 , 2 , 1 , 1
maka 7 bukan anggota dr suatu data sehingga nilainya benar

Translation (English):

ordered -> 45, 68, 72, 74, 76, 1, 2, 8,
position 1, 30, position 8, 68, position 21, 76, position 9, 2, 1, 1
therefore, 7 is not a member of the data set so the answer is correct

Figure 2. Response of S17 to Problem 1


urut dari bawah keatas dengan cara yg terkecil hingga ke besar
5 , 8 , 10 , 15 atau 20 , 15 atau 70 , 8 atau 80 , 5 , 0 atau 0
posisi 1 adalah angka 5 dan benar

Translation (English):

Arrange from the bottom to the top in ascending order from the smallest to the largest.

5, 8, 10, 15 or 20, 15 or 70, 8 or 80, 5, 0 or 0

The position 1 is the number 5 and it is correct.

Figure 3. *Response of S10 to Problem 1*

S26 provided the most explicit written representation of the initial problem information. The response documented the relevant numerical information, determined the total number of observations (36 students), and identified the value with the highest frequency (70, occurring 15 times). The response therefore showed a clear written connection between the information extracted from the task and the statistical quantities subsequently reported. S17 also used relevant information from the task to determine the required statistical quantities. The response indicated the number of observations and identified the value associated with the highest frequency. However, these elements were documented more concisely than in S26, with less explicit separation between the information provided by the task and the information being sought.

S10 presented the required numerical information more directly. The response stated that there were 36 students and identified 70 as the most frequent value, with a frequency of 15 students. Although the required statistical result was obtained, the response did not explicitly document the known information and the quantities being sought as separate elements before presenting the result.

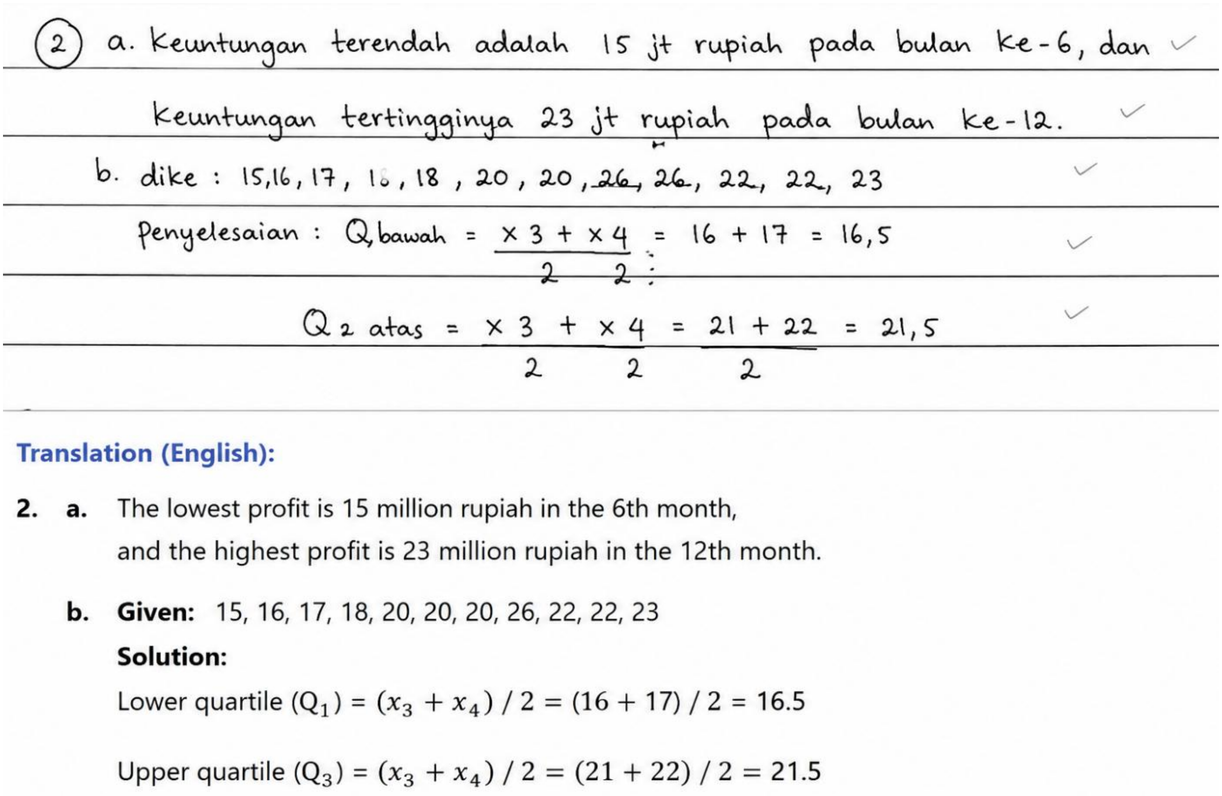
Table 4. Evidence of Students' Performance at the Understanding-the-Problem Stage

Participant	Relevant information	Required information	Statistical evidence in response	Characteristic of written evidence
S26	Explicitly documented	Explicitly represented	36 observations; mode = 70; frequency = 15	Explicit and structured
S17	Evident from the response	Evident but concise	Number of observations and highest-frequency value identified	Present but concise
S10	Reflected through the reported result	Not explicitly documented as a separate element	36 students; mode = 70; frequency = 15	Limited explicit documentation

As shown in Table 4, all three responses contained evidence that relevant statistical information was used, but the explicitness and organization of the problem-understanding process differed. S26 documented the initial information and its relation to the required quantities most explicitly, whereas S17 represented the relevant information more concisely. S10 produced the required numerical result but provided less explicit written evidence of the initial identification of what was given and what was being sought. The comparison therefore indicates a distinction between obtaining the required result and explicitly documenting the problem-understanding process. In the responses examined, correct statistical results were not consistently accompanied by the same level of explicit evidence showing how the problem was initially interpreted.

Devising a Plan

The *devising a plan* stage concerns how students establish a mathematical strategy for addressing the task before carrying out the required procedures. In the responses examined, both students used an appropriate approach for determining the lower and upper quartiles, but the strategy was represented with different degrees of explicitness in their written work.


2 a. Keuntungan terendah adalah 15 jt rupiah pada bulan ke-6, dan
Keuntungan tertingginya 23 jt rupiah pada bulan ke-12.
b. dike : 15,16,17, 18, 18, 20, 20, ~~26~~, 26, 22, 22, 23
Penyelesaian : Q bawah = $\frac{x3}{2} + \frac{x4}{2}$ = 16 + 17 = 16,5
Q 2 atas = $\frac{x3}{2} + \frac{x4}{2} = \frac{21 + 22}{2}$ = 21,5

Translation (English):

2. a. The lowest profit is 15 million rupiah in the 6th month, and the highest profit is 23 million rupiah in the 12th month.
b. **Given:** 15, 16, 17, 18, 20, 20, 20, 26, 22, 22, 23
Solution:
Lower quartile (Q_1) = $(x_3 + x_4)$ / 2 = (16 + 17) / 2 = 16.5
Upper quartile (Q_3) = $(x_3 + x_4)$ / 2 = (21 + 22) / 2 = 21.5

Figure 4. Response of S26 to Problem 2

2 diket : tabel keuntungan restoran setiap bulan dalam 1 tahun ✓
dit : a. berapakah keuntungan terendah dan tertinggi ✓
b. hitunglah nilai kuartil atas dan kuartil bawah ✓
dari keuntungan restoran
peny : * keuntungan terendah didapat ke bulan 6 yaitu 15 jt
tertinggi — " — 12 yaitu 23 jt ✓
* 15 , 16 , 16 , 17 , 18 , 20 , 20 , 20 , 21 , 22 , 22 , 23 ✓
jumlah data = 12. ✓
- kuartil bawah : 15, 16, 16, 17, 18, 20 — $\frac{16 + 17}{2}$ = 16,5 jt // 2
- kuartil atas :
20, 20, 21, 22, 22, 23 — $\frac{21 + 22}{3}$ = 21,5 jt //

Translation (English):

2. **Given:** Table of the restaurant's profit every month in 1 year.
Asked: **a.** What are the lowest and highest profits?
b. Calculate the upper quartile and the lower quartile of the restaurant's profit.
Solution:
* The lowest profit is in month 6, which is 15 million rupiah;
the highest is in month 12, which is 23 million rupiah.
* 15, 16, 16, 17, 18, 20, 20, 20, 21, 22, 22, 23
Number of data = 12.
- Lower quartile:
15, 16, 16, 17, 18, 20 → (16 + 17) / 2 = 16.5 million rupiah
- Upper quartile:
20, 20, 21, 22, 22, 23 → (21 + 22) / 2 = 21.5 million rupiah

Figure 5. Response of S32 to Problem 2

In S26's response, the student first represented the data and then proceeded to calculate the quartiles. The response identifies the data set, then uses the quartile procedure to determine the lower and upper quartiles. For the lower quartile, the student used the relevant middle values and calculated $(16 + 17)/2 = 16.5$, while the upper quartile was obtained from the corresponding upper-half values as $(21 + 22)/2 = 21.5$. Thus, the response contains evidence of an appropriate mathematical procedure for determining the two quartiles. However, the strategy itself was not explicitly stated as a separate planning step; it was primarily embedded in the subsequent organisation of the data and calculations. S32 demonstrated a more explicitly documented sequence before completing the quartile calculations. The response first stated the information provided and what was required, then organized the observations and separated the data into the lower and upper portions before determining the corresponding quartile values. The student subsequently used the middle values of each portion to obtain $Q_1 = 16.5$and $Q_3 = 21.5$. Compared with S26, therefore, S32 made more of the intermediate organisation underlying the selected procedure visible in the written response.

Table 5. Evidence of Students' Performance in Devising a Plan

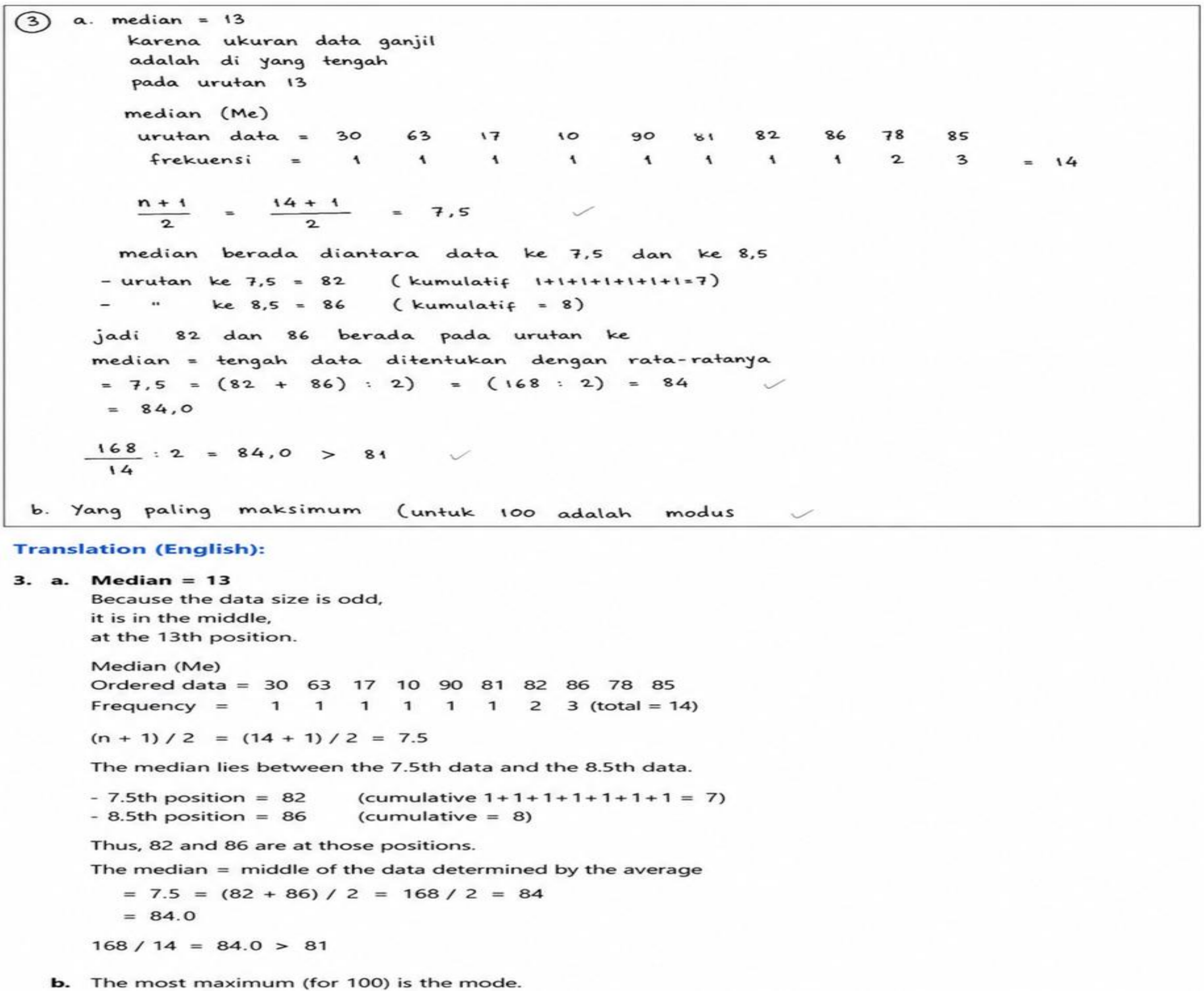
Participant	Strategy Evidence	Data Organization	Procedure Selected	Explicitness of Planning
S26	Evident through the sequence of data representation and quartile calculation	Data represented before calculation	Appropriate quartile procedure	Implicit / embedded in procedure

S32	Explicit sequence from given information to data organization and quartile calculation	Data organized and divided into lower and upper portions	Appropriate quartile procedure	More explicitly documented

As shown in Table 5, both responses contained evidence of an appropriate strategy for solving the quartile task, but the planning process was represented differently. S26 demonstrated the selected strategy primarily through the mathematical procedure itself, whereas S32 made more of the organisational steps preceding the calculation explicit. The distinction therefore concerns not the appropriateness of the strategy, but the extent to which the planning process was externalised in the written solution. This comparison indicates that evidence of *devising a plan* may appear through different forms of written activity. A student may demonstrate an appropriate plan implicitly through an ordered mathematical procedure, while another may make the intermediate organisational decisions more explicit before carrying out the calculation.

Carrying Out the Plan

The *carrying out the plan* stage concerns students' implementation of the selected strategy through mathematically appropriate and consistent procedures. The responses examined in Figures 6 and 7 showed different difficulties in translating the intended statistical procedure into a valid calculation.


3 a. median = 13
karena ukuran data ganjil
adalah di yang tengah
pada urutan 13

median (Me)
urutan data = 30 63 17 10 90 81 82 86 78 85
frekuensi = 1 1 1 1 1 1 1 1 2 3 = 14

$\frac{n+1}{2} = \frac{14+1}{2} = 7,5$

median berada diantara data ke 7,5 dan ke 8,5
- urutan ke 7,5 = 82 (kumulatif 1+1+1+1+1+1+1=7)
- " ke 8,5 = 86 (kumulatif = 8)

jadi 82 dan 86 berada pada urutan ke
median = tengah data ditentukan dengan rata-ratanya
= 7,5 = (82 + 86) : 2) = (168 : 2) = 84
= 84,0

$\frac{168}{14}$: 2 = 84,0 > 81

b. Yang paling maksimum (untuk 100 adalah modus

Translation (English):

3. a. Median = 13
Because the data size is odd,
it is in the middle,
at the 13th position.

Median (Me)
Ordered data = 30 63 17 10 90 81 82 86 78 85
Frequency = 1 1 1 1 1 1 2 3 (total = 14)

(n + 1) / 2 = (14 + 1) / 2 = 7.5

The median lies between the 7.5th data and the 8.5th data.

- 7.5th position = 82 (cumulative 1+1+1+1+1+1+1 = 7)
- 8.5th position = 86 (cumulative = 8)

Thus, 82 and 86 are at those positions.
The median = middle of the data determined by the average
= 7.5 = (82 + 86) / 2 = 168 / 2 = 84
= 84.0

168 / 14 = 84.0 > 81

b. The most maximum (for 100) is the mode.

Figure 6. Response of S17 to Problem 3

S17 attempted to determine the median from the given data by first identifying the number of observations and locating the positions around the centre of the distribution. The response calculated $(14+1)/2 = 7.5$ and subsequently identified the 7.5th and 8.5th positions as 82 and 86, respectively. The student then calculated their average as 84. Although this sequence shows an attempt to apply a positional procedure for determining the median, the written work does not show a consistent ordering of the data before identifying the positions. In addition, the subsequent calculation comparing $168/14 = 84\ with\ 81\ does\ not\ provide$ a mathematically meaningful verification of the median. Thus, the difficulty in S17's response was not limited to arithmetic accuracy; it involved inconsistency between the organization of the data, identification of positional values, and subsequent calculation.

3. Urutan sepatu — Banyak sepatu

Urutan sepatu	Banyak sepatu
38	10
39	10
40	10
41	10
mean < 42	30
median < 43	10
Modus : 43 ✓	40.

dll.

mean $(38\times10)+(39\times10)+(40\times10)+(41\times10)+(42\times30)+(43\times10)+(44\times40)$
$(10+10+10+10+30+10+40)$
$= 380+390+400+410+1260+430+1760 = 5030$
$= 9340$
$= \frac{9340}{220} = 42,45$ dibulatkan menjadi 42

median : ~~dompus~~ data ke 76 dan 110 (jumlah 110)
39 : 10 (1 - 10)
40 : 10 (11 - 20)
41 : 10 (21 - 30)
42 : 30 (31 - 60)
43 : 10 (61 - 70)
44 : 40 (71 - 110)
median tengah = ~~43~~ 42 - 43

Modus : 43.

b. mean < 42 , median = 43 42 ~~46~~, modus = 43

modus : 43 Paling laku keras

moda : 43

~~43~~ yg dapat disarankan stok bungaknya alasan 43

Translation (English):

3. Shoe sizes — Number of shoes

Shoe sizes	Number of shoes
38	10
39	10
40	10
41	10
mean < 42	30
median < 43	10
Mode = 43 ✓	10

etc.

- Mean

Mean $= (38\times10)+(39\times10)+(40\times10)+(41\times10)+(42\times30)+(43\times10)+(44\times40)$
$(10+10+10+10+30+10+40)$
$= 380+400+410+1260+430+1760 = 5030$
$= 5030 / 110 = 45.18$ rounded to 42

- Median

Median: data number 76 and 110 (total 110)
39 : 10 (1 – 10)
40 : 10 (11 – 20)
41 : 10 (21 – 30)
42 : 30 (31 – 60)
43 : 10 (61 – 70)
44 : 40 (71 – 110)
Middle median: between 42 and 43

- Mode

Mode = 43.

b. Mean < 42, median = 42, mode = 43

Mode: 43 is the most in demand (sells the most).

Conclusion: 43.

The size 43 is recommended to be stocked in the largest amount because size 43 is the most demanded.

Figure 7. Response of S10 to Problem 3

S10 attempted to determine the mean, median, and mode from a frequency distribution of shoe sizes. For the mean, the response multiplied each shoe size by its corresponding frequency and obtained a sum of 5030. However, the subsequent denominator and arithmetic are inconsistent with the frequency distribution presented. The listed frequencies total 120, whereas the response uses a different denominator in the calculation. The response also reports a rounded mean of 42 without a consistent preceding calculation. A similar issue appears in the determination of the median. S10 constructed cumulative frequency intervals for the shoe sizes, but the identified central positions do not correspond consistently to the total frequency of the distribution. The response ultimately identifies values around 42 and 43 as the middle, but the positional reasoning shown does not support this conclusion. For the mode, the response identifies 43, although the frequency distribution shown assigns the highest frequency to shoe size 44. Consequently, S10's response shows an attempted use of appropriate statistical procedures, but the execution is not consistently aligned with the data and the definitions of the measures being calculated.

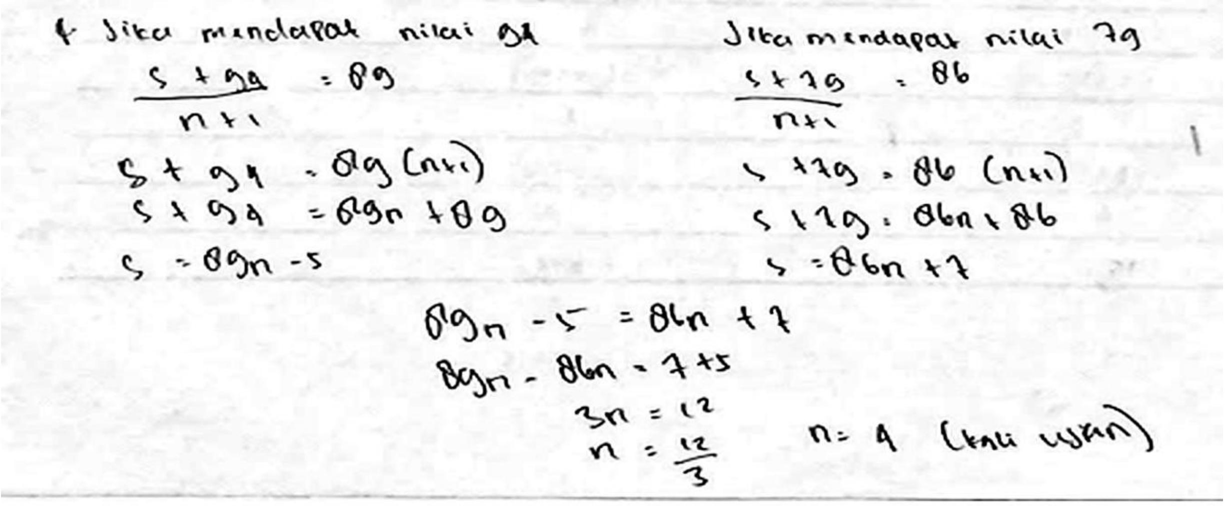
Table 6. Evidence of Students' Performance in Carrying Out the Plan

Participant	Procedure Attempted	Evidence in Written Response	Main Issue	Stage-Level Pattern
S17	Median based on positional values	Calculates $(n + 1)/2 = 7.5$, identifies two central positions, and averages the corresponding values	Data ordering and positional identification are not consistently established; subsequent calculation is not mathematically coherent	Procedure attempted but not consistently executed
S10	Mean, median, and mode from a frequency distribution	Uses score × frequency for the mean and cumulative frequencies for the median	Frequency total, denominator, central positions, and mode identification are inconsistent with the displayed distribution	Multiple procedural inconsistencies despite use of relevant procedures

The comparison indicates that both responses contained evidence of students attempting to implement recognisable statistical procedures, but neither response demonstrated fully consistent execution across the calculations examined. S17's difficulty was concentrated in the connection between data ordering, positional reasoning, and the resulting median calculation, whereas S10 showed inconsistencies across several components of the statistical calculation, including the mean, median, and mode. The findings therefore show that identifying an appropriate procedure and carrying it out consistently are distinguishable aspects of problem-solving performance.

Looking Back

The *looking back* stage concerns students' examination and evaluation of a solution after completing the mathematical procedures. The analysis therefore examined whether students' written responses contained explicit evidence of checking the obtained result, reconsidering the procedure, or relating the result back to the original problem.



4 Jika mendapat nilai 94

$$\frac{S+94}{n+1} = 89$$

$S + 94 = 89(n+1)$

$S + 94 = 89n + 89$

$S = 89n - 5$

Jika mendapat nilai 79

$$\frac{S+79}{n+1} = 86$$

$S + 79 = 86(n+1)$

$S + 79 = 86n + 86$

$S = 86n + 7$

$89n - 5 = 86n + 7$

$89n - 86n = 7 + 5$

$3n = 12$

$n = \frac{12}{3}$ $n = 4$ (kali ujian)

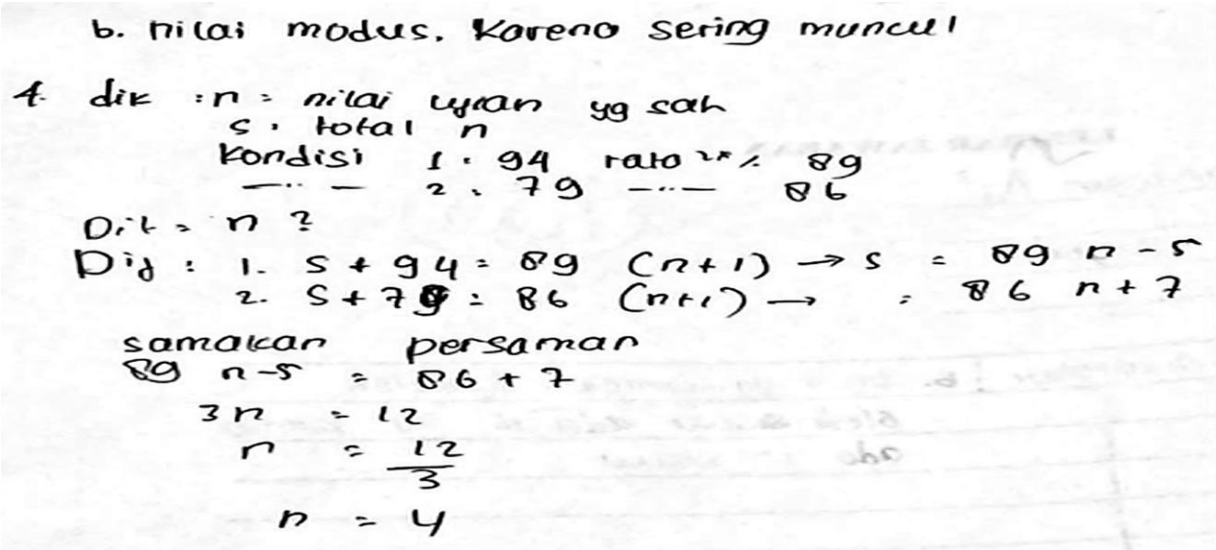
Translation (English):

4. If a score of 94 is obtained:
   (S + 94) / (n + 1) = 89
   S + 94 = 89(n + 1)
   S + 94 = 89n + 89
   S = 89n − 5

   If a score of 79 is obtained:
   (S + 79) / (n + 1) = 86
   S + 79 = 86(n + 1)
   S + 79 = 86n + 86
   S = 86n + 7

   Therefore:
   89n − 5 = 86n + 7
   89n − 86n = 7 + 5
   3n = 12
   n = 12/3 = 4 (four students)

Figure 8. Response of S17 to Problem 4


b. nilai modus. Karena sering muncul

4. dik : n = nilai ujian yg sah
s = total n
Kondisi 1. 94 rata² 89
— — 2. 79 — — 86
Dit = n ?
Dij : 1. S + 94 = 89 (n+1) → S = 89 n − 5
2. S + 79 = 86 (n+1) → = 86 n + 7
samakan persaman
89 n − 5 = 86 + 7
3n = 12
n = 12/3
n = 4

Translation (English):

b. The mode value, because it occurs frequently.

4. Given:
n = number of students who took the exam
S = total score

Condition 1: if a score of 94 is obtained, the average is 89.
Condition 2: if a score of 79 is obtained, the average is 86.

Asked: n?

Solution:
1. $S + 94 = 89(n + 1)$
$S = 89n - 5$

2. $S + 79 = 86(n + 1)$
$S = 86n + 7$

Equating the two equations:
$89n - 5 = 86n + 7$
$3n = 12$
$n = 12/3$
$n = 4$

Figure 9. Response of S22 to Problem 4

In the response shown in Figure 8, S17 completed the mathematical calculation and arrived at a numerical result. However, the written response did not document a subsequent step in which the result was checked against the original problem or the preceding calculation. Similarly, S22 completed the calculation shown in Figure 9 but did not explicitly verify the result or provide a written evaluation connecting it to the problem context. The two responses therefore showed procedural completion without explicit documentation of a subsequent checking or evaluation process. This pattern was observed in both S17, with an overall score of 62.5, and S22, with an overall score of 100. The same pattern across these two achievement levels indicates that the overall test score did not, by itself, determine whether explicit looking-back evidence appeared in the written response.

Cross-Stage Pattern of Mathematical Problem-Solving Performance

The responses examined across the four Pólya stages did not show the same pattern of performance. Differences were apparent in both the way students documented their work and the consistency with which they carried out the procedures. In the *understanding the problem* stage, S26 explicitly identified the relevant information, while S17 presented the same type of information more briefly and S10 moved directly to the numerical result without separately stating what was given and what was required. In *devising a plan*, S26 and S32 used an appropriate procedure for determining the quartiles, although S32 made the intermediate organization of the data more visible in the written response. The *carrying out the plan* responses showed another form of variation. S17 attempted a positional procedure for determining the median but did not maintain a consistent connection between the ordering of the data and the identified positions. S10 also used recognizable procedures for the mean, median, and mode, but several parts of the calculation were inconsistent with the frequency distribution presented. In the *looking back* responses, both S17 and S22 completed the calculations without documenting a subsequent check or evaluation of the results.

These responses point to a distinction between the presence of a problem-solving procedure and the extent to which the process was made visible and maintained consistently in the written work. Students could identify relevant information and use familiar statistical procedures, but these aspects were not necessarily accompanied by equally explicit planning, consistent execution, or verification. The differences were therefore observable not only in the final answers but also in the intermediate work through which students approached the statistical tasks. The stage-level evidence also did not correspond consistently with the students' overall achievement classifications. S10, for example, received an overall score of 50, yet his response to the task examined contained several identifiable statistical procedures, even though their execution was not fully consistent with the displayed data. S22 received an overall score of 100, but the response examined did not contain explicit evidence of checking or evaluating the solution. S17, with an overall score of 62.5, likewise showed different characteristics across the tasks examined: one response identified relevant problem information, while another showed difficulties in executing the procedure and documented no subsequent verification. Thus, the overall score provides information about students' aggregate performance, but it does not necessarily describe the particular characteristics of their work at each Pólya stage.

Overall, the cross-stage evidence shows that students' problem-solving processes varied within and across the four stages of Pólya's framework. The variation was evident in the explicit representation of problem information, the visibility of planning, the consistency of procedural execution, and the presence of verification. Among the responses examined, the *looking back* stage was notable for the absence of explicit checking or evaluation, while the earlier stages showed more visible evidence of information use and procedure application.

DISCUSSION

The students' responses show that solving a statistical problem involved more than selecting a correct operation. Across the four Pólya stages, the written work differed in how students represented the problem, organised a solution, carried out the calculation, and dealt with the result. This pattern is visible even among students who occupied different overall achievement categories. Previous studies using Pólya's framework have generally reported differences in students' abilities across the stages, particularly in understanding the problem, planning a solution, and carrying out procedures (Amalia & Salafudin, 2022; Kusumaningrum et al., 2023; Novitasari et al., 2022; Yunarti & Roesdiana, 2021). Similar variation has been reported in studies of junior secondary students working on different mathematical topics (Sakinah et al., 2024) and in studies examining problem solving in relation to students' mathematical characteristics (Pangaribuan et al., 2025). The present responses add a more specific observation: the differences were not confined to the final answer. They were also visible in the form and consistency of the work produced while students moved through the stages of solving a statistical task. This distinction matters because a student's written solution can show evidence of understanding or procedure use without showing the same level of explicitness at every subsequent stage.

The responses to the planning tasks illustrate this distinction clearly. S26 and S32 both selected an appropriate procedure for determining the quartiles, but they did not present the planning process in the same manner. S32 made the organization of the data before calculation more visible, whereas S26 moved more directly from the data to the quartile procedure. In statistical reasoning, planning is not simply a preliminary statement of intention; it involves deciding how to organize the information in the problem and which procedure is appropriate for answering the question (Chervany et al., 1980). Research on statistical problem solving likewise shows the value of examining students' actions during the solution process rather than restricting assessment to the final output (Woodard & Lee, 2021). The finding here is therefore not that S26 lacked a plan. His response provides evidence that the plan was embedded in the mathematical work rather than separately articulated. This distinction is also relevant to the broader use of Pólya's stages in mathematics education, where the stages provide an analytical structure for examining how students approach a problem rather than merely assigning a label to the completed answer (Amalia & Salafudin, 2022; Ully & Hakim, 2022). In the present data, planning was consequently observable in the organization of the solution, but its explicitness varied between students.

A different pattern appeared when students implemented their selected procedures. S17 attempted to determine the median through positional reasoning, but the relationship between the ordering of the data, the identified positions, and the subsequent calculation was not maintained consistently. S10 also used recognisable procedures for calculating the mean, median, and mode, yet several parts of the calculation did not consistently correspond to the frequency distribution presented in the response. The issue in these cases is therefore broader than a simple computational mistake. The students showed evidence of knowing which statistical procedures were relevant, but they did not always translate those procedures into a mathematically coherent sequence of operations. This distinction has also appeared in research examining students' answer processes, where errors can occur within the procedure even when a student has engaged with an appropriate solution approach (Sari et al., 2022). It is particularly relevant in statistical problem solving because students must coordinate data representation, procedural decisions, and numerical operations rather than perform an isolated calculation (Fauzi et al., 2025; Woodard & Lee, 2021). Process-oriented research offers a useful way to view this distinction because it attends to the sequence of actions through which a solution is produced (Ogut et al., 2024). Thus, the responses of S17 and S10 illustrate a gap between recognising or initiating an appropriate procedure and sustaining that procedure accurately throughout the solution.

The same distinction helps explain why the overall achievement scores did not always correspond to what was visible at an individual Pólya stage. S10 received an overall score of 50, but the response examined contained identifiable attempts to use procedures for the mean, median, and mode. However, the calculations themselves were not consistently aligned with the data. S22 obtained an overall score of 100, yet the response examined for the fourth problem contained no explicit checking or evaluation after the calculation had been completed. These cases do not invalidate the overall classification; rather, they show that the aggregate score and the evidence contained in a particular problem-solving stage provide different information. Process-oriented assessment has been developed because information about how students work through a task can be obscured when assessment focuses mainly on the outcome (Hao et al., 2025; Jeltova et al., 2007). Evidence from mathematical problem-solving research also indicates that different components of solving can be associated with different cognitive demands (Pyke et al., 2017), while metacognitive monitoring has been linked to differences in mathematics achievement (Cohors-Fresenborg et al., 2010). The present data provide a concrete classroom-level illustration of this issue: a high overall score does not ensure that every component of the written problem-solving process is explicitly demonstrated, just as a lower overall score does not mean that every stage is performed poorly.

The clearest limitation in the written responses concerned *looking back*. S17 and S22 completed the calculations but did not record a subsequent action that checked the result, reconsidered the procedure, or evaluated whether the answer was reasonable in relation to the problem. This absence matters because verification requires students to return to the solution after it appears complete and examine the relationship between the result, the procedure, and the original problem. Research on verification strategies has shown that students do not necessarily engage in such checking spontaneously and that verification itself involves distinguishable forms of mathematical activity (Mashiach-Eizenberg & Zaslavsky, 2004). Work on metacognition similarly places monitoring and regulation within mathematical problem solving rather than treating them as activities separate from solving (Jacobse & Harskamp, 2012; Tachie, 2019). This finding is also consistent with evidence linking metacognitive monitoring to mathematics achievement, although achievement level alone does not determine whether monitoring will be visible in a particular response (Cohors-Fresenborg et al., 2010). In the present study, the key finding is more limited and more directly supported by the data: students who completed the required calculations did not make a verification process visible in their written responses. The absence of such documentation should not be interpreted as proof that no internal checking occurred, but it does show that verification was not part of the observable written solution.

Taken together, the evidence provides a more differentiated picture of students' statistical problem solving than the overall achievement categories alone. The responses contained evidence of problem interpretation, planning, and procedure use, but these processes varied in explicitness and were not always carried through consistently to verification. This pattern aligns with work that treats mathematical problem solving as an activity composed of interconnected processes rather than a single performance outcome (Liljedahl & Cai, 2021; Kilpatrick, 2016). It also accords with research examining students' actions during statistical problem solving, where the organisation and sequence of activity provide information about how a solution is constructed (Fauzi et al., 2025; Woodard & Lee, 2021; Ogut et al., 2024). The Pólya framework is useful here not because it simply divides students into four stages, but because it makes differences in students' written activity visible: one student may make the problem information explicit, another may embed planning within the procedure, another may begin an appropriate calculation but lose consistency during execution, and another may complete the calculation without documenting verification. The resulting picture is therefore not a simple hierarchy from high to low ability. It is a pattern of variation across the processes through which students construct and evaluate statistical solutions. This process-level view is particularly relevant to didactical mathematics because it keeps attention on the mathematical activity that becomes visible in students' work, while avoiding the stronger claim that a written response fully captures students' internal reasoning.

CONCLUSIONS

This study examined Grade VIII students' statistical problem solving through Pólya's four stages: understanding the problem, devising a plan, carrying out the plan, and looking back. Analysis of the students' written responses showed that performance varied across these stages and was not consistently represented by overall achievement scores. The responses provided evidence of students identifying relevant information and applying statistical procedures, but the explicitness of problem interpretation and planning varied, while procedural execution was not always consistent. The most apparent difficulty was found in the *looking-back* stage. The responses examined did not contain explicit evidence of verifying the obtained results or evaluating their reasonableness after completing the calculations, including the response of a student classified in the highest achievement category. The findings indicate that completing a statistical procedure and obtaining a result do not necessarily demonstrate that the student has documented the full problem-solving process. Accordingly, examining students' performance at the level of individual Pólya stages provides a more differentiated account of their statistical problem solving than an overall achievement score alone. For didactical practice, assessment of statistical problem solving should therefore attend to students' evidence of understanding, planning, execution, and verification, with particular attention to making post-solution checking an explicit component of students' mathematical work.

ADDITIONAL INFORMATION

Section	Description
Funding	This research received no external funding.
Acknowledgment	The author would like to express sincere gratitude to the mathematics teachers, students, and administration staff of SMKN 1 Karawang for their support and participation during the research process. Appreciation is also extended to the validators and lecturers who provided valuable suggestions and feedback for the improvement of the developed learning media.
Conflict of Interest	The author declares no conflict of interest.
Data Availability	The data supporting the findings of this study are available from the corresponding author upon reasonable request.
Author Contributions	The author was responsible for conceptualisation, research design, data collection, media development, data analysis, manuscript writing, and final approval of the submitted version.

REFERENCES

Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences, 25*(2), 203–239. https://doi.org/10.1080/10508406.2016.1143370

Aini, N. N., Rosyidi, A. H., & Hasnawati. (2023). Kemampuan pemecahan masalah siswa berdasarkan teori Polya pada pembelajaran *problem-based learning* materi statistika. *Jurnal Math-UMB.EDU, 11*(1), 28–41. https://doi.org/10.36085/mathumbedu.v11i1.5650

Anwar, R. B., & Rahmawati, D. (2017). Symbolic and verbal representation process of student in solving mathematics problem based Polya's stages. *International Education Studies, 10*(10), 20–28. https://doi.org/10.5539/ies.v10n10p20

Chan, M. C. E., & Clarke, D. (2017). Structured affordances in the use of open-ended tasks to facilitate collaborative problem solving. *ZDM Mathematics Education, 49*, 951–963. https://doi.org/10.1007/s11858-017-0876-2

García, T., Rodríguez, C., González-Castro, P., González-Pienda, J. A., & Torrance, M. (2016). Elementary students' metacognitive processes and post-performance calibration on mathematical problem-solving tasks. *Metacognition and Learning, 11*, 139–170. https://doi.org/10.1007/s11409-015-9139-1

Güner, P., & Erbay, H. N. (2021). Metacognitive skills and problem-solving. *International Journal of Research in Education and Science, 7*(3), 715–734. https://doi.org/10.46328/ijres.1594

Hedges, S., Harkness, S. S., & Given, K. (2022). Attending to middle school students' understanding of study design: The often overlooked or contrived step in the statistical problem-solving process. *Middle Grades Research Journal, 13*(1), 53–65. https://doi.org/10.1108/MGRJ-05-2022-0004

Hidayanto, E., & Rahmatina, D. (2020). Characteristics of student statistical reasoning in mathematical problem solving. *AIP Conference Proceedings, 2215*, 030028. https://doi.org/10.1063/5.0003653

Kania, N., & Ratnawulan, N. (2022). Kompetensi matematika: Kemampuan pemecahan masalah matematis siswa menurut Polya. *Journal of Research in Science and Mathematics Education, 1*(1), 17–26. https://doi.org/10.56855/jrsme.v1i1.10

Koichu, B., Parasha, R., & Tabach, M. (2021). Who-is-right tasks as a means for supporting collective looking-back practices. *ZDM–Mathematics Education, 53*, 831–846. https://doi.org/10.1007/s11858-021-01264-z

Lee, S. Y. (2016). Students' use of "look back" strategies in multiple solution methods. *International Journal of Science and Mathematics Education, 14*, 701–717. https://doi.org/10.1007/s10763-014-9599-9

Liljedahl, P., & Cai, J. (2021). Empirical research on problem solving and problem posing: A look at the state of the art. *ZDM–Mathematics Education, 53*, 723–735. https://doi.org/10.1007/s11858-021-01291-w

Luthfiyah, N. I., Haryanto, H., & Firmansyah, F. (2022). Analisis kemampuan pemecahan masalah matematika berdasarkan teori Polya pada siswa kelas VIII SMP IT Insan Mulia Manokwari. *THEOREMA: The Journal Education of Mathematics, 2*(2), 15–22. https://doi.org/10.36232/theorema.v2i2.1731

Makwakwa, E. G., Mogari, D., & Ogbonnaya, U. I. (2024). First-year undergraduate students' statistical problem-solving skills. *Teaching Statistics, 46*(1), 8–23. https://doi.org/10.1111/test.12359

Meisy, M. S. M., Aniah, T., & Salamah, S. (2022). Analisis kemampuan pemecahan masalah materi penyajian data kelas IX SMP Sungai Raya. *Jurnal Cartesian (Jurnal Pendidikan Matematika), 1*(2), 91–99. https://doi.org/10.33752/cartesian.v1i2.2515

Özcan, Z. Ç., & Eren Gümüş, A. (2019). A modeling study to explain mathematical problem-solving performance through metacognition, self-efficacy, motivation, and anxiety. *Australian Journal of Education, 63*(1), 116–134. https://doi.org/10.1177/0004944119840073

Rosidah, R., Budayasa, I. K., & Juniati, D. (2018). An analysis of statistical reasoning process of high school students in solving the statistical problem. *Journal of Physics: Conference Series, 1028*(1), 012125. https://doi.org/10.1088/1742-6596/1028/1/012125

Villamin, P., Lopez, V., Thapa, D. K., & Cleary, M. (2025). A worked example of qualitative descriptive design: A step-by-step guide for novice and early career researchers. *Journal of Advanced Nursing, 81*(8), 5181–5195. https://doi.org/10.1111/jan.16481

Amalia, N. R., & Salafudin. (2022). Penerapan pendekatan *problem solving* versi Polya untuk meningkatkan kemampuan pemecahan masalah matematika siswa MTsS Hifal Pekalongan. *CIRCLE: Jurnal Pendidikan Matematika, 2*(1), 70–77. https://doi.org/10.28918/circle.v2i01.5094

Chervany, N. L., Benson, P. G., & Iyer, R. K. (1980). The planning stage in statistical reasoning. *The American Statistician, 34*(4), 222–226. https://doi.org/10.1080/00031305.1980.10483032

Cohors-Fresenborg, E., Kramer, S., Pundsack, F., & others. (2010). The role of metacognitive monitoring in explaining differences in mathematics achievement. *ZDM–Mathematics Education, 42*, 231–244. https://doi.org/10.1007/s11858-010-0237-x

Fauzi, I. S., Sutrisno, S., Selviana, V., & Bakhrudin, B. (2025). Analysis of vocational students' problem-solving competencies in resolving case studies in statistics: Analysis of vocational students' problem-solving competencies. *Journal of Vocational Education Studies, 8*(1), 58–76. https://doi.org/10.12928/joves.v8i1.11995

Hao, S., Pan, H., & Zhang, D. (2025). A process-oriented approach to assessing high school students' mathematical problem-solving competence: Insights from multidimensional eye-tracking analysis. *Education Sciences, 15*(6), 761. https://doi.org/10.3390/educsci15060761

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2007). Dynamic assessment as a process-oriented assessment in educational settings. *Advances in Speech Language Pathology, 9*(4), 273–285. https://doi.org/10.1080/14417040701460390

Jacobse, A. E., & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning, 7*, 133–149. https://doi.org/10.1007/s11409-012-9088-x

Kilpatrick, J. (2016). Reformulating: Approaching mathematical problem solving as inquiry. In P. Felmer, E. Pehkonen, & J. Kilpatrick (Eds.), *Posing and solving mathematical problems* (pp. 113–127). Springer. https://doi.org/10.1007/978-3-319-28023-3_5

Kusumaningrum, S. R., Furqon, M. N., Ulya, W. J. U., & Amaliyah, F. (2023). Analisis kemampuan pemecahan masalah matematika dalam menyelesaikan soal cerita berdasarkan prosedur Polya. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam, 2*(2), 162–169. https://doi.org/10.55606/jurrimipa.v2i2.1539

Mashiach-Eizenberg, M., & Zaslavsky, O. (2004). Students' verification strategies for combinatorial problems. *Mathematical Thinking and Learning, 6*(1), 15–36. https://doi.org/10.1207/s15327833mtl0601_2

Novitasari, J., Pujiastuti, H., & Sudiana, R. (2022). Analisis kemampuan pemecahan masalah siswa menurut teori Polya dalam menyelesaikan soal cerita matematika. *Wilangan: Jurnal Inovasi dan Riset Pendidikan Matematika, 3*(3), 231–235. https://doi.org/10.56704/jirpm.v3i3.13402

Ogut, B., Webb, B., Hicks, J., Circi, R., & Yin, M. (2024). Exploring mathematical problem-solving through process mining: Insights from large-scale assessment log data. *Computers in the Schools*, 1–31. https://doi.org/10.1080/07380569.2024.2416422

Pangaribuan, F., Sidabutar, R., & Darmayanti, R. (2025). Exploring high school students' 3D geometry problem-solving: Role of cognitive style and mathematical ability. *Al-Jabar: Jurnal Pendidikan Matematika, 16*(1), 363–376. https://doi.org/10.24042/ajpm.v16i1.24361

Pyke, A. A., Fincham, J. M., & Anderson, J. R. (2017). When math operations have visuospatial meanings versus purely symbolic definitions: Which solving stages and brain regions are affected? *NeuroImage, 153*, 319–335. https://doi.org/10.1016/j.neuroimage.2017.03.046

Sakinah, N. I. N., Sari, A. A., Septiana, N. P. G. R., Herman, T., & Hasanah, A. (2024). Kemampuan pemecahan masalah matematis siswa pada materi relasi dan fungsi kelas VIII pada salah satu SMP Negeri di Bandung. *Eksponen, 14*(1), 1–11. https://doi.org/10.47637/eksponen.v14i1.975

Sari, N., Saragih, S., Napitupulu, E. E., Rakiyah, S., Suciawati, H., & Anim, A. (2022). Analysis of answering process on problem solving ability through problem-based learning model. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 7*(2), 269–287. https://doi.org/10.31943/mathline.v7i2.287

Tachie, S. A. (2019). Metacognitive skills and strategies application: How this helps learners in mathematics problem-solving. *Eurasia Journal of Mathematics, Science and Technology Education, 15*(5), Article em1702. https://doi.org/10.29333/ejmste/105364

Ully, A. C., & Hakim, D. L. (2022). Kemampuan pemecahan masalah matematis siswa dengan tahapan Polya. *Didactical Mathematics, 4*(1), 156–162. https://doi.org/10.31949/dm.v4i1.2014

Woodard, V., & Lee, H. (2021). How students use statistical computing in problem solving. *Journal of Statistics and Data Science Education, 29*(sup1), S145–S156. https://doi.org/10.1080/10691898.2020.1847007

Yunarti, Y., & Roesdiana, L. (2021). Analisis kemampuan peserta didik menyelesaikan soal cerita berdasarkan indikator Polya. *Jurnal Ilmiah Soulmath: Jurnal Edukasi Pendidikan Matematika, 9*(2), 129–142. https://doi.org/10.25139/smj.v9i2.3405