ISSN 2355-4185(p), 2548-8546(e)
DOI: https://doi.org/10.24815/jdm.v12i2.48250

Student Learning Processes on the Measure of Central Tendency using the Theory of Didactical Situations

Wulan Putri Utami[1], Mohamad Gilar Jatisunda[1*]

[1]Mathematics Education Department, Universitas Majalengka, Indonesia
*Email: *g.jatisunda@unma.ac.id*

Received: 26 July 2025; Revised: 26 October 2025; Accepted: 30 October 2025

Abstract. *In the current data-driven society, developing a deep understanding of statistical concepts is essential for students to interpret and communicate quantitative information effectively. This study explores how students construct a sense of the measures of central tendency—mean, median, and mode—through interactions within didactical situations designed according to the Theory of Didactical Situations (TDS). Using an exploratory case study, sixteen tenth-grade senior high school students participated in two instructional sessions structured around five TDS phases: devolution, action, formulation, validation, and institutionalisation. Data were collected through observations, video recordings, written work, and field notes, and analyzed using deductive content analysis guided by the TDS framework. The findings reveal that learning structured around TDS phases fosters the emergence of intuitive strategies and promotes conceptual understanding of mean, median, and mode. Students gradually shifted from procedural computation toward reflective reasoning and contextual interpretation. This study contributes to mathematics education by demonstrating how TDS-based instruction can foster epistemic autonomy and conceptual growth in statistical learning.*

Keywords*: measures of central tendency, student exploration, theory of didactical situations*

How to cite: Utami, W. P., & Jatisunda, M. G. (2025). Student Learning Processes on the Measure of Central Tendency using the Theory of Didactical Situations. *Jurnal Didaktik Matematika, 12* (2), 384-406. DOI: https://doi.org/10.24815/jdm.v12i2.48250

Introduction

Statistical literacy has increasingly been recognized as a critical competence for individuals navigating the complexities of the 21st century. In an era of abundant data, interpreting, evaluating, and communicating statistical information is essential for academic achievement, informed citizenship, and professional success (Hoffrén, 2021; Wallman, 1993). As Carlson (2002) has argued, statistical competencies constitute a fundamental foundation for preparing workers to meet the evolving demands of data-driven workplaces. This view is supported by educational scholarship that positions statistical literacy as a key civic skill, enabling individuals to critically engage with quantitative claims across diverse contexts such as health, finance, media, and environmental issues (Lipič & Ovsenik, 2020; Schield, 2005). Recent research also emphasises the need for pre-service teachers to develop strong statistical reasoning skills to teach effectively and model data-informed thinking in schools (Forgasz et al., 2024). Hence, statistical literacy is increasingly regarded as a central component of contemporary education systems worldwide.

Within school mathematics, measures of central tendency, namely, mean, median, and mode, are fundamental topics in introductory statistics. These concepts are typically introduced in primary education and revisited throughout secondary schooling, serving as key tools for understanding data distributions and characterising datasets. However, research suggests that educational materials such as textbooks often present these concepts fragmentedly, focusing mainly on procedural skills rather than conceptual understanding (Parra-Fica et al., 2024) . Consequently, opportunities for fostering deep comprehension from an early stage may be limited.

Despite their early introduction, many students struggle to grasp the conceptual meanings of these statistical measures and their appropriate applications. Studies conducted in diverse educational settings, including rural and under-resourced schools, report that learners frequently struggle to interpret these measures beyond mere calculation (Karaca & Ay, 2025; Saidi & Siew, 2018). For example, Landtblom and Sumpter (2025). Found that although sixth-grade students could accurately compute mean and median values, they had difficulty discerning which measure best represented a given dataset. Semiotic conflicts compound such difficulties when students encounter symbolic or graphical data representations (Bogas et al., 2019). These findings highlight the necessity of pedagogical approaches integrating conceptual clarity with fluency in representations, thereby supporting critical reasoning about data beyond procedural computation.

Numerous studies have shown that students often prioritise procedural fluency over conceptual understanding when learning measures of central tendency. This limited perspective creates difficulties in interpreting the meaning and appropriate use of these measures, underscoring the need for learning environments that integrate conceptual reasoning with procedural competence. Recent studies have echoed these findings, revealing that even with modern curricular reforms, many students continue to rely on procedural strategies rather than conceptual reasoning. For example, Ismail and Chan (2015) found that misconceptions about central tendency persist despite years of formal instruction, indicating unresolved conceptual gaps. Similarly, Maryati and Priatna (2018) reported that such misconceptions remain prevalent among pre-service teachers, thereby hindering their statistical reasoning. Meanwhile, Meletiou-Mavrotheris and Lee (2024) demonstrated that statistical reasoning remains fragile without explicit conceptual scaffolding. Collectively, these studies emphasize the continued need to revisit classical findings through updated empirical contexts and pedagogical frameworks.

These patterns are also apparent within the Indonesian educational context. Dewi et al. (2024) identified epistemological and cognitive barriers among prospective teachers in understanding statistical concepts, particularly measures of central tendency. At the school level, Safitri and Ruli (2025) reported that vocational high school students exhibited low statistical reasoning skills, linked to a disconnect between mathematical procedures and meaningful real-world contexts. The shift to online learning during the COVID-19 pandemic further intensified these learning challenges, as students struggled to interpret datasets meaningfully through asynchronous digital instruction (Romadoni et al., 2023). Similarly, Fitri et al. (2023) documented limited statistical literacy among senior high school students in Aceh, revealing a tendency toward mechanical calculation without internalising representational or interpretive aspects. Together, these studies underscore the urgent need for instructional designs that prioritise sense-making over rote calculation, guiding learners to develop contextual understanding and supporting critical reasoning about measures beyond procedural computation.

The Theory of Didactical Situations (TDS), initially developed by Brousseau, highlights the vital role of the milieu—the interactive learning environment constituted by tasks, tools, peers, and feedback—in fostering autonomous student knowledge construction. Central to TDS are didactic situations in which learners engage independently with the milieu, constructing mathematical knowledge through active problem-solving rather than through direct teacher intervention. This theoretical framework has effectively promoted deep conceptual understanding across various mathematical domains (Hortelano & Prudente, 2024; Mathé & Mithalal, 2025). Notably, TDS has been successfully applied to interdisciplinary topics, including cryptography (Bartzia et al., 2024), computer programming (Abdüsselam et al., 2022), and arithmetic concepts such as factors and multiples (Sulistyani et al., 2024), where learner-milieu interaction supports meaningful mathematical activity. However, the relationship between TDS and the development of critical statistical reasoning remains underexplored. As noted by Ben-Zvi and Makar (2016) Critical statistical reasoning requires learners to question the validity, context, and interpretation of data, which aligns with the epistemic autonomy central to TDS. Moreover, recent work by Özmantar and Zeynep (2023). Demonstrated that TDS-based learning environments foster reflective reasoning by positioning students as active constructors of knowledge, suggesting their potential for enhancing statistical literacy and reasoning in data-rich contexts.

In the context of statistics education, particularly regarding measures of central tendency, TDS offers a promising framework for moving beyond procedural instruction towards situated conceptual knowledge construction. When immersed in well-designed didactical situations, students are encouraged to explore, hypothesise, and negotiate meanings collaboratively, with the milieu comprising real datasets, manipulatives, visualisations, and problem contexts that demand interpretation and justification of statistical measures. Empirical studies have demonstrated that such approaches foster statistical reasoning and support the emergence of conceptual understanding alongside procedural fluency, particularly when the didactic contract facilitates a-didactical engagement (Yenil et al., 2023).

Although various studies have examined students' understanding of measures of central tendency, most have focused on procedural competence rather than how learners autonomously construct statistical meaning through structured interactions in a didactical milieu. Few studies have explored the application of the TDS in statistics, particularly in the context of Indonesian senior high school education. Thus, there exists a research gap in understanding how TDS can support students' conceptual growth in statistics by linking procedural fluency with contextual reasoning and reflective justification. This study contributes to filling that gap by analysing how senior high school students construct meanings of mean, median, and mode through active engagement in a TDS-based learning milieu. By situating this inquiry within Indonesia's curriculum context, the study extends the international discourse on didactical theories to local pedagogical challenges (Forgasz et al., 2024).

Method

This study employed a qualitative, exploratory case study design. The qualitative approach was chosen as it is suitable for exploring, explaining, and interpreting learning phenomena in natural contexts, allowing for an in-depth and multi-dimensional understanding of the complex process of constructing the concept of measures of central tendency (Creswell & Creswell, 2017). The case study focused on teaching and learning central tendency measures—mean, median, and mode—within a specific and dynamic context structured by the principles of the TDS (Yin, 2014). The study aimed to explore and report the complex interactions among three key elements: students, learning tasks, and the learning milieu.

The study involved sixteen senior high school students (aged 16–17) from a suburban public school in Indonesia who had not yet received formal instruction on measures of central tendency. Students worked in groups of four during two instructional sessions. In addition, five students (S1–S5) were purposively selected for short follow-up interviews to capture a range of emerging reasoning patterns identified in classroom observations and written work.

Data were collected through three primary sources: participant observations, video recordings of group discussions, and students' written work. Observation notes captured students' behaviours and interactions across the five phases of the didactical situation, while video recordings provided detailed evidence of their reasoning processes. Students' written responses to five contextual tasks, designed to elicit exploration of mean, median, and mode, were used to document individual and group reasoning. All tasks were piloted with two external student groups and reviewed by two mathematics education experts to ensure clarity and alignment with the TDS phases. To support the analysis, the five contextual tasks administered during the learning sessions are summarised in Table 1.

Table 1. Contextual tasks used in the study

No.	Task Title	Data Description	Questions
1	Defective Shoes Data	Number of defective shoes recorded over 40 days: 2, 1, 0, 1, 3, 2, 4, 2, 1, 0, 2, 1, 1, 5, 1, 0, 1, 2, 3, 1, 0, 4, 1, 0, 3, 4, 0, 1, 1, 2, 1, 0, 1, 1, 0, 3, 2, 0, 1, 0	a. Construct a frequency table. b. Compute the mean, mode, and median. c. Draw a bar chart.
2	Raka's Phone Usage	Daily screen-time duration (minutes) over 8 days: 120, 135, 140, 130, 125, 600, 150, 145	a. Compute the mean, median, and mode. b. Identify the most appropriate measure of central tendency and justify. c. Explain whether the data represent Raka's typical behaviour.
3	Mathematics Quiz Scores	Quiz scores of 15 students: 70, 75, 80, 75, 85, 70, 90, 75, 80, 75, 85, 90, 70, 75, 80	a. Construct a frequency table. b. Determine the mode. c. Explain whether the most frequent score represents the class's average performance.
4	Dina's Daily Steps	Number of steps recorded over 10 days: 3200, 4500, 5200, 4900, 3000, 3100, 7000, 4800, 5000, 5200	a. Draw a line graph. b. Compute the mean and median. c. Describe the activity pattern.
5	Lala's Sleep Duration	Sleep duration (hours) over 14 days: 6, 7, 6, 8, 7, 6, 12, 11, 7, 6, 7, 8, 7, 6	a. Compute the mean, median, and mode. b. Identify the most appropriate measure of central tendency. d. Justify your answer based on the data pattern.

The study was carried out over two instructional sessions organised according to the five phases of the TDS. Prior to implementation, the researchers designed and piloted five contextual tasks to ensure their clarity and alignment with TDS principles. During instruction, students engaged with the tasks through the devolution, action, formulation, validation, and institutionalization phases, with the teacher deliberately withholding procedural guidance to maintain an a-didactical milieu. Video recordings, field notes, and students' written work were collected throughout the sessions to capture their emerging strategies, representations, and reasoning. Following the classroom implementation, five focal students (S1–S5) were purposely selected for semi-structured interviews based on variations in their observed reasoning patterns. These interviews provided more profound insight into students' decision-making processes across the TDS phases and strengthened triangulation, confirming interpretations drawn from classroom observations and artefacts.

Data were collected through multiple qualitative sources to ensure depth and credibility. Field notes from participant observations documented verbal and non-verbal interactions across the five TDS phases. At the same time, video recordings of group discussions were transcribed verbatim to capture students' reasoning and collaborative discourse. Students' written work, including problem-solving sheets and reflective responses, provided additional evidence of their conceptual development. Follow-up interviews with five focal students (S1–S5) were also conducted to clarify key reasoning patterns observed during the sessions. Trustworthiness was strengthened through strategies recommended in qualitative research. (Creswell & Creswell, 2017; Miles et al., 2018), including methodological triangulation, peer debriefing, member checking, and establishing inter-coder agreement, with two coders reaching 95% consensus. These measures ensured credibility, dependability, and confirmability in the analysis, consistent with standards outlined by Lincoln and Guba (1985) and operationalised by Nowell et al. (2017).

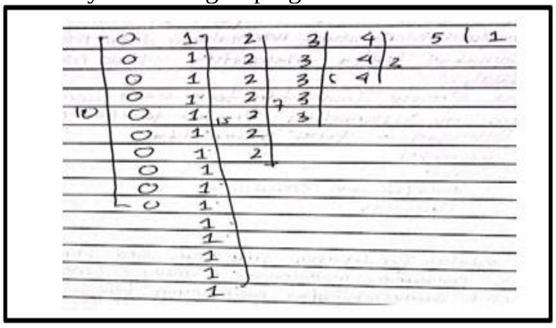
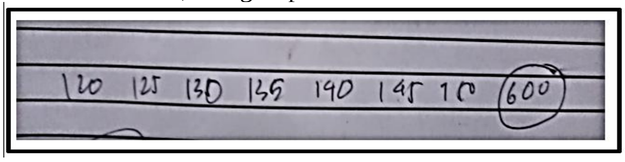
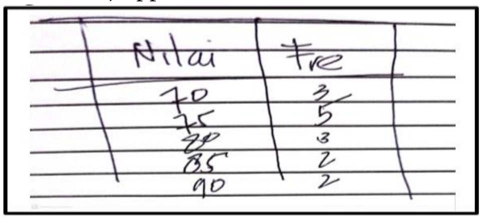
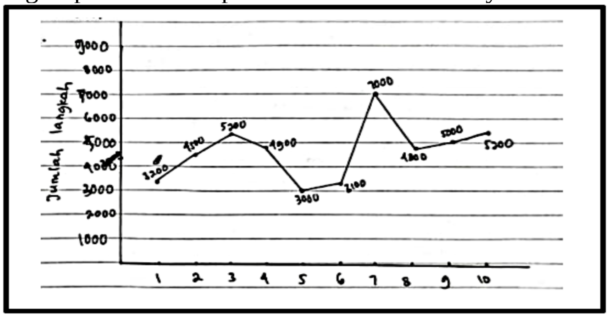
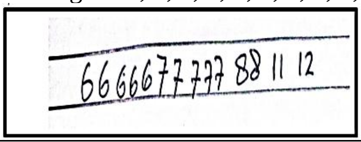
Data were analysed using a deductive qualitative content analysis approach as outlined by Braun and Clarke (2006), with the five phases of the TDS serving as the a priori analytical framework. All video-recorded group discussions were transcribed verbatim and subjected to initial coding according to the TDS phases—devolution, action, formulation, validation, and institutionalisation. Codes were refined through iterative review of the recordings, enabling verification of interpretive accuracy and identification of emergent nuances in students' reasoning. Field notes were cross-checked systematically with transcript data to ensure coherence across data sources. Students' written work was subsequently analysed to corroborate and enrich the interpretive themes, thereby enhancing analytical depth and completeness. To ensure analytic rigor, multiple strategies to assess trustworthiness were employed. Member checking enabled participants to verify the accuracy of the interpreted interactions. At the same time, inter-coder reliability was established through collaboration with an external mathematics education expert with over fifteen years of qualitative research experience. Coding discrepancies were resolved through iterative consensus-building until complete agreement was achieved, consistent with the reliability standards proposed by Miles et al. (2018).

Results and Discussion

Devolution Phase: Mobilising the Situation

In the devolution phase, the teacher introduced five contextual problems without providing formulas, procedural steps, or guidance related to measures of central tendency. All tasks were open-ended, allowing students to select the initial strategies they considered most appropriate. Through this approach, the teacher transferred the responsibility of reasoning to the students and created an a-didactical milieu—an instructional situation in which learners rely on their intuition, prior experiences, and self-constructed ways of interpreting. This situation is presented in Table 2.

Table 2. Students' response activities

<table>
<tr><th>No.</th><th>Problems</th><th>Students' Responses Within the Group</th></tr>
<tr><td>1</td><td>Defective Shoe Data</td><td>Each group has successfully identified the data by listing all defective items vertically and then grouping the identical numbers.
</td></tr>
<tr><td>2</td><td>Raka's Mobile Phone Usage</td><td>The data identification is shown by listing the sequence of phone-usage duration values. However, one group circled the number 600.
</td></tr>
<tr><td>3</td><td>Mathematics Quiz Scores</td><td>Each group understands how to create a simple table containing the columns "Value" and "Frequency." They counted how many times each value (70, 75, 80, 85, and 90) appeared.

<table><tr><th>Score</th><th>Frequency</th></tr><tr><td>70</td><td>3</td></tr><tr><td>75</td><td>5</td></tr><tr><td>80</td><td>3</td></tr><tr><td>85</td><td>2</td></tr><tr><td>90</td><td>2</td></tr></table></td></tr>
<tr><td>4</td><td>Dina's Daily Step Count</td><td>A group draws a simple line chart with the day axis and the step-count axis.
</td></tr>
<tr><td>5</td><td>Lala's Sleep Duration</td><td>Each group's identification has led them to arrange the data from the smallest to the largest: 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 11, 12.
</td></tr>
</table>

The students' work across the five tasks demonstrated that they relied on their pre-existing knowledge before receiving any formal instruction. In the defective-shoes task, students created vertical columns containing stacks of the numbers 0, 1, 2, 3, 4, and 5. This was a manual grouping method used to identify the most frequently occurring value. They selected this strategy independently, without teacher guidance, indicating that the milieu enabled them to construct meaning through intuition. As stated by student (S3): *"So we can see which one appears the most. It is easier to look at it this way."* When asked why they arranged the numbers in this manner, one student responded:

In the task on Raka's mobile phone usage, students redrew the data as a sequence of numbers and quickly noticed the presence of an extreme value. The value 600 was immediately circled, indicating that the students perceived it as unusual compared to the rest of the dataset. During the interview, one student (S5) explained: *"All the numbers are close to each other except 600. That one makes the result go way up if you add and divide it."* This response suggests that the student had begun to associate the impact of an outlier with changes in the mean value.

This response illustrates early awareness of outliers, along with an intuitive interpretation of the mean as the "fairest" value because it incorporates all data points. In the mathematics quiz task, students independently constructed a frequency table. They counted the occurrences of each score, noting that 75 appeared 5 times. One student (S1) concluded: *"Seventy-five comes up the most, so it is the most common. It is kind of like the average."*

In Dina's daily-step task, students drew on intuitive understandings of habits and normal ranges when interpreting the data. They identified values between 3,000 and 5,000 as "normal," while the two values of 7,000 were immediately recognized as unusually high. During an interview, one student (S4) explained this as presented in the following interview transcript: one explained:

Researcher	:	*"Look at this data. Do you see any numbers that appear different or stand out?"*
Student (S4)	:	*"Yes, the 7,000 ones look different. The others are all around three to five thousand."*
Researcher	:	*"Why do you think the 7,000 values feel different?"*
Student (S4)	:	*"Maybe that was a day when she walked a lot more. They are far from the others."*
Researcher	:	*"If you had to estimate Dina's average steps without using a formula, what would you guess?"*
Student (S4)	:	*"Probably around four thousand. Most of the numbers are around that. The 7,000 only shows up sometimes."*

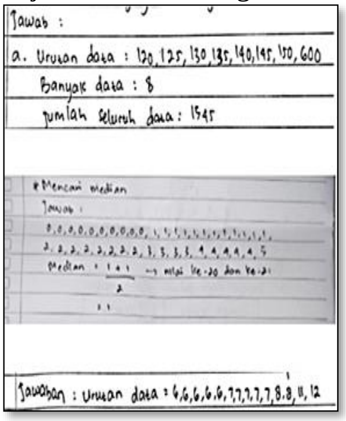
These comments show that students relied on rational strategies without using formal calculations. They estimated the average based on clusters of values rather than the arithmetic mean. The line plots they drew reflected how they visualized the daily fluctuations to make sense of the data. In Lala's sleep-duration task, students employed a similar approach. They listed the data sequentially, then grouped the values at 6 and 7 hours because they appeared most frequently. The two values of 11 and 12 hours received special attention as unusual observations. One student

(S2) noted: *"The 11 and 12 are like when she was exhausted. Usually, she sleeps only 6 or 7 hours. So the ones that appear the most show her usual sleep pattern."*

The devolution phase across these tasks demonstrated that the five activities successfully created learning conditions that promoted student autonomy and knowledge construction. Through these open-ended tasks, students worked independently, formed initial understandings of the situations, and interpreted data using their everyday intuition and logical reasoning. Their spontaneous interactions with the problems led to the emergence of proto-concepts related to measures of central tendency, which served as the foundation for a deeper statistical understanding in subsequent phases. These findings align with the core principles of the TDS, which emphasises that knowledge is not transmitted directly by the teacher but is constructed by learners through active engagement with tasks and the designed milieu.

Action Phase: Engaging with the Milieu

In the action phase, students engaged directly with the milieu—comprising the dataset, the problem context, and minimal instructions—without any formal procedures provided by the teacher. Their exploration relied on the intuitive strategies developed during the devolution phase. At this point, they began making decisions, testing strategies, refining their thinking, and experiencing cognitive conflict when their expectations did not align with the data's structure. The milieu was intentionally designed to create epistemic tension through extreme values, asymmetric distributions, and varied frequencies, prompting students to recognise the need to select appropriate measures of central tendency. This phase marked the emergence of autonomous learning. The students' written responses in Figure 6 demonstrate how these intuitive strategies played out in practice, offering clear evidence of their decision-making, adjustments, and cognitive efforts during the action phase. The students' written responses in Figure 1 illustrate how these intuitive strategies unfolded in practice, providing concrete evidence of the decisions, adjustments, and cognitive work they performed during the action phase.



Answer:
a. Ordered data: 120, 125, 130, 135, 140, 145, 150, 600
Number of data points: 8
Total of all data: 1545

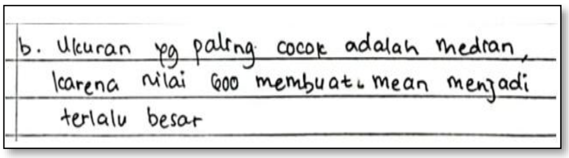
Finding the Median
Answer:
0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,4,4,4,4,4,5
Median= $\frac{(1+1)}{2} = 1$ (values in the 20th and 21st positios)

Answer: Orderred data: 6,6,6,6,6,7,7,7,7,7,8,8,11,12

Figure 1. Students ordering the data

The students' worksheets in Figure 1 show that they systematically ordered the data, identified extreme values, and then determined an appropriate measure of central tendency. In the *"Finding the Median"* section, they arranged the numbers in ascending order. They marked the two central values, demonstrating a visual and intuitive method for identifying the "middle value" and constructing early conceptual understanding consistent with the TDS. At this stage, students began to employ more structured strategies beyond intuition. They organised the data, produced simple frequency counts, highlighted outliers, examined value intervals, and performed initial mean calculations to detect emerging patterns. These behaviours indicate the use of representational strategies, such as ordering and clustering, even though their reasoning had not yet been fully formalised. This developing sophistication is reflected in the comment of one student (S2): *"When the data are ordered, it becomes clear which values are unusual and which ones are normal. That way, we can decide which value is representative."*

Then, productive epistemic tension emerged when students discovered that their calculations did not align with their initial expectations. This was particularly evident in their work on the fourth task in Figure 2, where they computed the mean of the mobile phone usage data and obtained an average that was significantly higher than most of the other values.

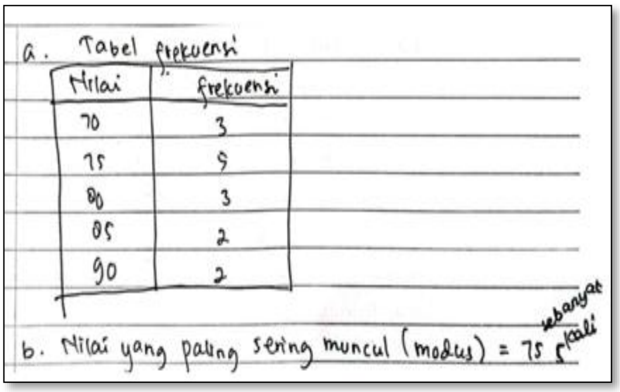

b. Ukuran yg paling cocok adalah median, karena nilai 600 membuat mean menjadi terlalu besar

The most appropriate measure is the median, because the value of 600 makes the mean become too large

Figure 2. Student notes showing data summation and the comment "too large" referring to the mean result

Students noted that "the value 600 makes the mean too large," indicating their awareness that an extreme value was affecting the representativeness of the mean. This awareness is consistent with statements made during the interview, such as *"Why did it become bigger? The others are only around 300–400,"* and *"There's one value that's really high, and that's what makes the average stranger."* Such realisations show that students were beginning to understand that the mean is not always an appropriate measure when a dataset contains outliers—an important conceptual insight that seldom develops without the presence of cognitive conflict.

In the action phase, it became evident that several students had begun independently comparing different measures of central tendency, particularly the mean and the mode. This was especially clear in their work on the second task in Figure 3.


a. Tabel frekuensi

Nilai	frekuensi
70	3
75	5
80	3
85	2
90	2

b. Nilai yang paling sering muncul (modus) = 75 sebanyak 5 kali

a. Frequency Table

Score	Frequency
70	3
75	5
80	3
85	2
90	2

b. The most frequently occurring score (mode) is 75, appearing five times.

Figure 3. Two students are discussing while pointing to the most frequent value in the frequency table

The students noted that although the value 75 was indeed the mode because it appeared most frequently, it did not necessarily represent the overall performance of the class. The students then calculated a mean of 78.33 and identified it as a more representative measure for describing the learning outcomes of all students. This observation aligns with interview comments student (S4): *"Mode appears the most, but does it represent everything? Maybe not."*

This indicates that the students were beginning to recognise that each measure of central tendency serves a different purpose and responds differently to the shape of the data distribution. Their work shows that they were no longer relying solely on procedural computation but had begun to evaluate their results conceptually. Throughout the action phase, the teacher offered no corrections or direct answers. This is evident from the students' worksheets, where they independently ordered the data, constructed diagrams, calculated the mean, median, and mode, and explained their reasoning without explicit guidance. In the task on Raka's mobile phone usage, for example, students noted that "the value 600 makes the mean too large," a conclusion derived from their own reading of the data rather than teacher intervention. This demonstrates that the milieu, rather than the teacher, functioned as the primary source of feedback, consistent with the principles of the TDS, which emphasise that feedback should arise from the situation itself. This shift also reflects a change in the didactical contract: students stopped looking to the teacher for confirmation and instead sought strategies aligned with the characteristics of the data. Field notes and worksheet evidence show that students increasingly consulted the data or their peers rather than the teacher, further reinforcing this shift.

Formulation Phase: Structuring and Articulating Mathematical Ideas

In the formulation phase, students began to shift from merely carrying out exploratory actions toward articulating, organising, and communicating their mathematical ideas explicitly. At this stage, they started to connect actions with underlying reasons, such as explaining why the median may provide a more representative value or under what conditions the mode becomes relevant. The milieu continued to play an important role, yet students progressively developed more structured representations and explanations that could be understood, tested, and negotiated with their peers. The formulation phase marks the emergence of conditional reasoning, the ability to explain that measures of central tendency are not used mechanically, but must be selected in accordance with the characteristics of the data distribution. Students began to use explanatory language, visual sketches, tables, and simple arguments to convey their ideas. To illustrate the nature of students' work during the formulation phase, the following descriptive example can be used as Figure 4.

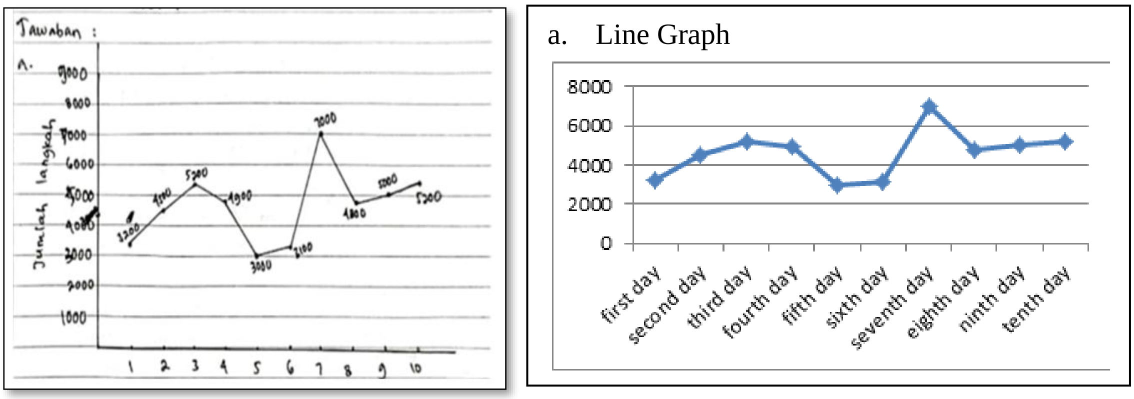
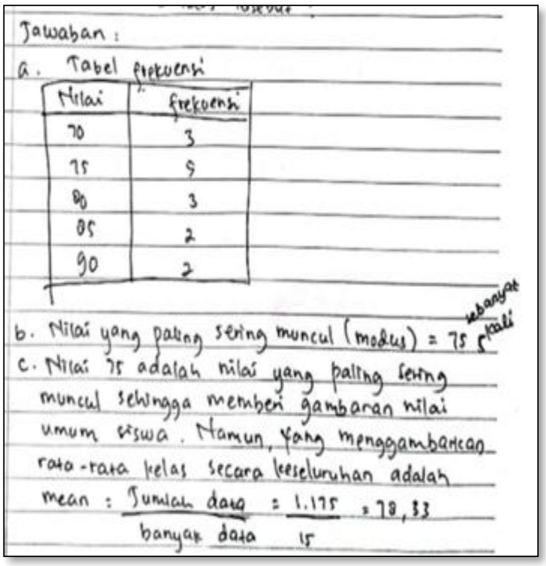


Figure 4. Number line sketch constructed by a student

At this stage, students displayed a clear transition from executing procedures to explaining why a particular statistical measure was appropriate for the dataset. Rather than treating the mean, median, and mode as mechanical outputs, they began examining whether each measure aligned with the structure of the distribution they had constructed. This shift became particularly evident when students re-evaluated the impact of the 600-minute value. One student observed that *"the 600 is far from the others, so the mean goes up and it doesn't represent the data pattern"* (S1), demonstrating an emerging awareness of the sensitivity of the mean to extreme values. Another student (S2) elaborated that the median remained stable because *"the median stays in the cluster, so it gives a more accurate central value"* , indicating an early conceptual understanding of robustness. A third student (S3) further synthesised this reasoning, stating that *"the mean didn't make sense for this dataset because it didn't match the cluster, but the median still showed the typical value"* .

Together, these micro-level justifications show that students were no longer relying on procedural correctness alone; instead, they were beginning to reason conditionally about when and why a measure of central tendency represents a dataset appropriately. This reasoning marks the culmination of the formulation phase in the TDS, where learners construct conceptual links between procedures and the characteristics of the milieu. The increasing use of distribution-based arguments—particularly references to clusters, outliers, and representativeness—naturally set the stage for the subsequent validation phase, in which students would test, defend, and negotiate these claims with their peers.

In the formulation phase, students' use of representations as thinking tools became more evident. They moved beyond recording numerical results and constructed frequency tables, and markings of extreme values to organise their understanding as shown in Figure 5.


Jawaban :

a. Tabel frekuensi

Nilai	Frekuensi
70	3
75	5
80	3
85	2
90	2

b. Nilai yang paling sering muncul (modus) = 75 sebanyak 5 kali

c. Nilai 75 adalah nilai yang paling sering muncul sehingga memberi gambaran nilai umum siswa. Namun, yang menggambarkan rata-rata kelas secara keseluruhan adalah mean : $\frac{\text{Jumlah data}}{\text{banyak data}} = \frac{1.175}{15} = 78{,}33$

a. Frequency Table

Score	Frequency
70	3
75	5
80	3
85	2
90	2

b. The mode is 75, as it appears five times.

c. A score of 75 characterizes the most common outcome among the students and reflects a central pattern in the distribution. However, the mean is the measure that more accurately represents the overall average performance of the class. A full interpretation would therefore require considering both the mode and the mean.

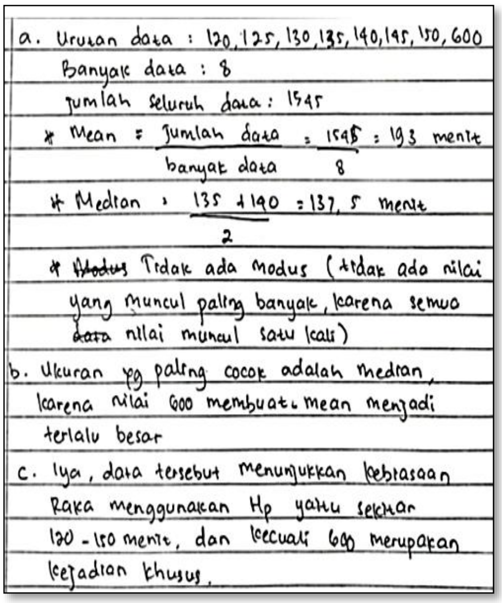
Figure 5. Student notes comparing the mean, median, and mode

In Figure 5, they arranged the distribution, compared the mode with the mean, and concluded that although the mode (75) appeared five times, it was not necessarily the most representative value, whereas the mean of 78.33 offered a more stable summary. Field notes also show that several students used number-line sketches to position the mean, median, and extreme values, indicating that visual representations served as cognitive scaffolds. During this phase, conditional reasoning began to emerge as students considered when each measure of central tendency was appropriate. In the mobile-phone-usage task, they explained that the mean was distorted by the extreme value of 600, while the median better reflected the typical pattern, demonstrating an early understanding of median robustness. Likewise, in the frequency-table task, students recognised that the mode is meaningful only when its frequency is dominant and contextually relevant.

Group collaboration provided opportunities for students to clarify and negotiate their arguments. Through discussion, they refined criteria for using the mean, median, and mode, shifting from viewing these measures as mechanical formulas to treating them as analytical tools that must align with the characteristics of the distribution. At the same time, the didactical contract shifted from teacher-directed confirmation to a learning environment requiring justification and argumentation. The teacher maintained a facilitative role, prompting students with reflective questions rather than giving answers. This absence of direct instruction encouraged students to articulate reasoning, engage in peer dialogue, and justify their conclusions using observable data patterns. This shift shows that students increasingly sought validation through analysis and collaboration, reflecting the essence of the formulation phase in TDS.

Validation Phase: Negotiating Meaning and Validating Strategies

In the validation phase, students moved beyond formulating and expressing ideas toward testing, defending, and justifying their strategies in front of their peers. The central focus of this stage was peer validation rather than teacher evaluation. The classroom became a space of mathematical argumentation, where ideas were assessed based on their consistency with the data rather than the authority of the teacher. The teacher maintained a non-intervention stance, providing only clarifying questions that encouraged students to construct their own validation. Through this process, students learned that the correctness of a strategy depends not on the teacher's approval but on how well the strategy aligns with the characteristics of the data distribution. The interaction captured in Figure 6 illustrates how this validation process unfolded in practice.


a. Urutan data : 120, 125, 130, 135, 140, 145, 150, 600
Banyak data : 8
Jumlah seluruh data : 1545
* Mean = $\frac{\text{Jumlah data}}{\text{banyak data}} = \frac{1545}{8} = 193$ menit
* Median = $\frac{135 + 140}{2} = 137,5$ menit
* ~~Modus~~ Tidak ada modus (tidak ada nilai yang muncul paling banyak, karena semua ~~data~~ nilai muncul satu kali)

b. Ukuran yg paling cocok adalah median, karena nilai 600 membuat mean menjadi terlalu besar

c. Iya, data tersebut menunjukkan kebiasaan Raka menggunakan Hp yaitu sekitar 120 - 150 menit, dan kecuali 600 merupakan kejadian khusus.

a. Ordered data: 120, 125, 130, 135, 140, 145, 150, 600.
Number of observations: 8
Sum of observations: 1545 minutes
- Mean = 1545 ÷ 8 = 193 minutes
- Median = mean of the 4th and 5th values = (135 + 140) ÷ 2 = 137.5 minutes
- Mode = No mode is present because each value occurs exactly once.

b. The most appropriate measure is the median, because the value of 600 makes the mean become too large

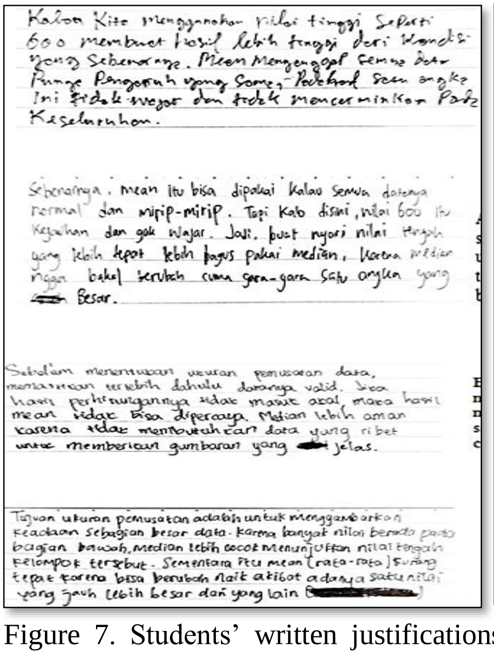
c. Yes, the data shows the typical duration of using a mobile phone which is around 120-150 minutes, and the 600 is an exceptional case

Figure 6. The object of discussion is whether the mean or the median

The written work presented in Figure 6 served as the central object of validation, providing claims that students subsequently examined for their consistency with the data. In this phase, the students did not simply present their computed values; instead, they scrutinized whether these results were adequate representations of the dataset. The inscription *"mean = 193 minutes"* immediately became a contested claim, as students identified the 600-minute value as an outlier that distorted the average. The note stating that *"600 makes the mean too large"* functioned as an initial critique, prompting students to test whether the mean remained a reliable measure under the influence of extreme values.

A second set of inscriptions proposed the median (137.5 minutes) as a more robust alternative, supported by written justifications such as *"median is more appropriate because 600 makes the mean too large"* and the classification of the 600-minute entry as *"a special case."* During validation, these claims were evaluated not on procedural grounds but on how well they aligned with the numerical cluster identified in the students' own notes (120–180 minutes). The emphasis on cluster stability and the identification of atypical values reflect an emergent process of argument-based validation, where students assessed each measure according to the characteristics of the distribution rather than its computational correctness.

The teacher's non-intervention stance further reinforced the validation dynamic. When students sought confirmation, the teacher redirected them with prompts such as *"What does the data tell you?"*, positioning the data, not the teacher, as the epistemic authority. This shift required students to justify their claims by explicitly referencing the distributional structure visible in their written work. Through this process, the students collectively validated the appropriateness of the median over the mean, not because the teacher endorsed it, but because the argument was most consistent with the empirical features of the dataset they had constructed. This analytic trajectory aligns closely with the additional written evidence presented in the next figure, which offers a more elaborate rationale for rejecting the mean and substantiates the students' preference for the median through explicit references to cluster patterns and the exceptional nature of the 600-minute value.


Kalau Kita menggunakan nilai tinggi Seperti 600 membuat hasil lebih tinggi dari Kondisi yang Sebenarnya. Mean Menganggap Semua Data Punya Pengaruh yang Sama, Padahal Satu angka Ini tidak wajar dan tidak mencerminkan Pada Keseluruhan.

Sebenarnya, mean itu bisa dipakai kalau semua datanya normal dan mirip-mirip. Tapi kalo disini, nilai 600 itu Kejauhan dan gak wajar. Jadi, buat nyari nilai tengah yang lebih tepat lebih bagus pakai median, karena median nggak bakal berubah cuma gara-gara satu angka yang Besar.

Sebelum menentukan ukuran pemusatan data, memastikan terlebih dahulu datanya valid. Jika hasil perhitungannya tidak masuk akal, maka hasil mean tidak bisa dipercaya. Median lebih aman karena tidak membutuhkan data yang ribet untuk memberikan gambaran yang jelas.

Tujuan ukuran pemusatan adalah untuk menggambarkan keadaan sebagian besar data. Karena banyak nilai berada pada bagian bawah, median lebih cocok menunjukkan nilai tengah kelompok tersebut. Sementara itu mean (rata-rata) kurang tepat karena bisa berubah naik akibat adanya satu nilai yang jauh lebih besar dari yang lain [redacted])

If we use a high value like 600, it makes the result apprear higher than is actually. The mean treats all data points as having the same influence, even though this one value is unusual and does not represent the overall data.
Actually, the mean can be used if all the data are normal and similar. But in this case, the value 600 is too far off and unusual. So, to find a more accurate central value, it's better to use the median because the median won't change just because of one large number.
Before choosing a measure of central tendency, we must first make sure the data are valid. If the calculation result doesn't make sense, then the mean cannot be trusted. The median is safer because it does not require complex or perfectly consistent data to give a clear picture.
The purpose of measures of central tendency is to describe the condition of most of the data. Since many of the values are in the lower range, the median is more suitable for representing the central value of that group. Meanwhile, the mean is less appropriate because it can increase due to the presence of a single value that is much larger than the others.

Figure 7. Students' written justifications evaluating the representativeness of the mean and median in the presence of an extreme value

Figure 7 presents a set of student inscriptions that functioned as the *object of validation* during the validation phase in TDS. At this stage, students no longer treated their calculations as final results but as claims requiring examination. The written statement "mean = 193 minutes" became a contested assertion once students identified the 600-minute entry as an outlier that distorted the average. Notes such as *"600 makes the mean go too high"* reflect an emerging process of critical scrutiny, where students evaluated the representativeness of statistical measures by testing them against the resistance of the milieu. Other inscriptions demonstrate how students validated the median as a more stable measure in the presence of an extreme value, an instance of *a posteriori justification* arising directly from their interaction with the dataset. By highlighting the numerical cluster of 120–180 minutes, students were not merely identifying patterns; they were evaluating the coherence between the median and the underlying distribution. This evidence shows a shift from procedural correctness to argument-based validation, a core characteristic of the validation phase in TDS.

Follow-up interviews further confirmed that students were indeed engaged in validation. S1 argued, *"The 600 is far from the others, so the mean goes up and it doesn't represent the data pattern,"* indicating a rejection of the mean based on inconsistency with the milieu. S2 emphasized that the median remained within the main cluster, *"so the median stays in that cluster and gives a more accurate central value"*, while S3 reasoned that *"the mean didn't make sense for this dataset because it didn't match the cluster."* These micro-justifications show that students validated or rejected strategies by referencing distributional structure, not procedural rules or teacher authority. The written work and interview excerpts show that students had fully entered the validation phase of TDS. Initially accepted strategies were challenged by the milieu, prompting students to refine their reasoning, justify their decisions with evidence, and converge on the median as the most appropriate measure. The figure thus illustrates how the milieu served as an epistemic validator, enabling students to construct mathematical truth through testing, rebuttal, and argumentation rather than procedural compliance.

Institutionalisation Phase: Bridging Personal Strategies with Formal Mathematical Knowledge

In the institutionalisation phase, the teacher resumes the role of epistemic authority to formalise the mathematical ideas that students previously constructed through interaction with the milieu. This phase does not function as a correction of students' answers; rather, it consolidates and names the strategies that emerged during validation, aligning them with established statistical concepts. At this point, students' informal arguments, such as recognising that the mean is sensitive to outliers, the median remains stable within the main cluster, and the mode is meaningful only when its frequency is contextually dominant, are reformulated into formal notions of mean sensitivity, median robustness, and context-dependent use of the mode. The teacher links these insights to the broader principle of selecting measures of central tendency according to the characteristics of the distribution. Through this institutionalising step, intuitive reasoning is transformed into academically recognised knowledge, enabling students to apply these concepts consistently across tasks and contexts.

Following the institutionalisation phase, the teacher formalised the informal strategies generated by the students by reorganising their emergent ideas into conventional statistical structures. The written work shown in Figure 8 illustrates this transition clearly: students' spontaneous descriptions of the mean, median, and mode—initially grounded in the distributional features of the dataset, were translated into formal definitions, advantages, and limitations. This transformation marks the moment when students' empirically derived reasoning is reframed into stable mathematical knowledge that aligns with institutional norms. The figure thus captures the consolidation of students' intuitive justifications into academically recognised properties of the measures of central tendency.

Ukuran Pemusatan	Definisi	Kelebihan	Keterbatasan ~~Ketergantungan~~
Mean	Nilai rata-rata dari seluruh data	Menggunakan semua data sehingga mewakili keseluruhan	Mudah terpengaruh jika ada nilai yang sangat tinggi atau sangat rendah
Median	Nilai tengah dari data yang sudah diurutkan	Tidak terpengaruh oleh nilai yang sangat tinggi atau sangat rendah	Tidak mempertimbangkan semua nilai, hanya fokus pada nilai tengah
Modus	Nilai yang paling sering muncul	Sangat mudah ditemukan	bisa memiliki lebih dari satu modus atau bahkan tidak memiliki modus

Measure of central tendency	Definition	Advantage	Limitation
Mean	The average value obtained by smming all data points and dividing by the number of data points.	Represents the entire dataset because it includes all values in the calculation	Highly sensitive to extreme values, which can distort the result
Median	The middle value of an ordered dataset	Not affected by extreme values	Does not take all data values into account, only the position of the middle value
Modus	The value that occurs most frequently in a dataset	Easy to determine	A dataset may have multiple modes or not mode at all, limiting its interpretive usefulness

Figure 8. Teacher-constructed summary of students' arguments and their institutionalised statistical forms

Figure 8 represents a pivotal moment within the institutionalisation phase as conceptualised in the TDS. In this phase, the teacher resumes the role of the *institutor*—the epistemic authority responsible for stabilising, naming, and formalising the knowledge objects that previously emerged through students' adidactical interaction with the milieu. Although the handwritten notes on the left side of the figure resemble student work, they are not spontaneous student productions; rather, they constitute the teacher's systematic recontextualisation of the arguments and justifications constructed by students during the formulation and validation phases. Empirical evidence of these contributions is seen in the interview data, in which S1 noted that *“the 600 makes the mean go up and it doesn’t represent the pattern,”* S2 explained that *“the median stays in the cluster and gives a more accurate value,”* and S3 concluded that *“the mean didn’t make sense because it didn’t match the cluster.”* These statements form the epistemic substrate from which the teacher assembled the conceptual table.

Through this reorganisation, the teacher enacted the central function of institutionalisation described by Brousseau: transforming “situational knowledge,” which is local, contextual, and tied to task-based experience, into “institutional knowledge,” which is generalised, stabilised, and transferable across contexts. Students’ interview statements—such as the observation that the mean “increases because of an extreme value”, were recast as the formal property of the mean’s sensitivity to outliers; the claim that the median “remains within the main cluster” was formalised into the principle of robustness; and the remark that the mode “is meaningful only when its frequency is dominant” was legitimised as a standard interpretative constraint in distributional analysis. In this way, the teacher elevated students’ emergent reasoning into a formal conceptual framework without erasing the epistemic pathways through which it was originally constructed.

The right-hand side of the figure functions as an objectification of knowledge, presenting explicit definitions, advantages, and limitations for each measure of central tendency. This marks the transition from argumentation embedded in the milieu to a conceptual system recognised by the statistical community. The convergence between students’ empirical reasoning and the institutionalised formulations underscores that institutionalisation does not overwrite students’ constructions; instead, it systematises and legitimises them within formal mathematical discourse. This process exemplifies the essence of institutionalisation in TDS: the teacher reanchors knowledge within the scientific domain, ensuring academic legitimacy while maintaining continuity with students’ adidactical learning experiences.

The findings of this study demonstrate a clear learning trajectory in which students progressed from intuitive, experience-based judgments toward increasingly structured and theoretically grounded reasoning about measures of central tendency. When examined through the lens of the TDS, the five observed phases collectively reveal how epistemic autonomy emerges through sustained interaction between learners and a deliberately designed milieu. This section critically interprets those findings and situates them in relation to prior scholarship.

A major insight emerging from the action phase is that students’ spontaneous strategies, such as ordering data, isolating extreme values, estimating central clusters, and constructing simple graphs, functioned as proto-representations. These strategies surfaced before students received any formal explanation, confirming Brousseau's (1997) assertion that a-didactical situations enable knowledge to arise from the dialectic between learner and milieu rather than teacher exposition. The productive cognitive conflicts that appeared, especially when students realised that the mean could be distorted by the value 600, illustrate what Barquero and Bosch (2015) describe as epistemic disequilibrium: a necessary condition for reorganising conceptual structures. This study enriches that theoretical claim by showing that disequilibrium did not merely trigger confusion, but rather initiated reflective comparison across the three measures of central tendency.

A critical pattern observed in the formulation phase was students' transition from computational accuracy to conditional reasoning. Students did not simply compute the mean, median, and mode but engaged in explicit argumentation about why certain measures were more representative than others, depending on the shape of the distribution. Their explanations, such as identifying the median as "more resistant to extreme values" or noting that the mode "only matters when its frequency is meaningful," demonstrate the emergence of inferential thinking. This aligns with Ben-Zvi and Makar's (2016) claim that representational diversity reinforces statistical reasoning, yet the present study adds nuance by showing that such reasoning arose even from non-standard, student-generated representations (e.g., hand-drawn number lines and self-constructed frequency tables). These findings highlight that representational agency—not the mere presence of representations—is essential for conceptual progress.

The validation phase provided strong evidence of a shift in the didactical contract. Students positioned the data, rather than the teacher, as the primary source of epistemic authority. Their arguments increasingly referenced distributional properties, such as clustering, skewness, and the influence of outliers. The debates between groups about whether the mean or median was more appropriate show a movement from procedural justification ("because that's the formula") toward data-driven justification ("because the cluster lies around…", "because the extreme value distorts…"). This supports Mathé and Mithalal's (2025) proposition that validation within TDS evolves into collective reasoning, where correctness is negotiated, not delivered. The present study refines this idea by showing that students not only validated each other's strategies but also revised their initial answers, suggesting a genuine epistemic shift rather than surface-level agreement.

The institutionalisation phase further revealed how students' informal language was transformed into disciplinary discourse. Through carefully timed teacher mediation, intuitive statements such as "the average is pulled up" or "this is the usual value" were formalised into the concepts of sensitivity to outliers, robustness, and representativeness. This resonates with Hortelano and Prudente's (2024) observation that institutionalisation is most effective when it builds directly on student-generated strategies rather than imposing abstract definitions. In this study, the teacher's consolidation enabled students to link experiential reasoning to canonical statistical structures, thereby strengthening their conceptual coherence and facilitating transferability to new datasets.

A broader theoretical contribution of this study lies in extending the epistemic reach of TDS into statistics education, a domain where conceptual understanding is inherently shaped by uncertainty, variability, and distributional reasoning rather than deterministic structures. While TDS has been predominantly applied in more rule-governed mathematical domains, the present findings indicate that carefully designed milieus can elicit productive shifts in learners' statistical reasoning as they engage with data. Recent empirical evidence has demonstrated that instructional designs emphasizing discourse, conceptual exploration, and authentic data analysis can significantly enhance students' progression from intuitive interpretations to more explicit, data-justified reasoning in statistics classrooms (Conway IV et al., 2019). Moreover, research employing think-aloud methods has shown that such environments activate real-time cognitive processes highly aligned with the interpretive nature of statistical thinking, enabling students to articulate and refine their understanding of representativeness, sampling variability, and causal interpretation (Reinhart et al., 2022). Collectively, these insights reinforce the theoretical compatibility between TDS and statistics education, suggesting that the milieu can be purposefully orchestrated to foreground core statistical ideas and support the development of reasoning pathways fundamental to statistical literacy.

Pedagogically, this research contributes a critical insight for Indonesian classrooms. Much of the literature identifies procedural teaching as a persistent obstacle to developing statistical reasoning (Dewi et al., 2024; Fitri et al., 2023; Safitri & Ruli, 2025). The present findings provide empirical confirmation of this issue but also demonstrate a viable alternative: positioning students as active interpreters of data through TDS-structured tasks. This aligns with the principles of the Merdeka Curriculum, which emphasises autonomy, reasoning, and reflective learning. The study, therefore, offers both theoretical and practical evidence that TDS-based instruction can help shift learning from procedural fluency toward conceptual depth.

Conclusion

This study aimed to investigate how students develop an understanding of mean, median, and mode through learning environments structured by the TDS. The findings demonstrate a clear progression from intuitive, experience-based interpretations to increasingly formal, data-driven, and context-sensitive reasoning as learners advanced through the five canonical TDS phases. Students' capacity to justify the appropriateness of statistical measures, negotiate interpretations collaboratively, and articulate the influence of distributional features underscores the pedagogical value of carefully designed a-didactical milieus. The study contributes to the broader field of mathematics and statistics education by extending the application of TDS to introductory statistical concepts—a domain in which the framework has received limited empirical attention. The results highlight the potential of TDS to cultivate epistemic agency, strengthen representational reasoning, and foster deeper engagement with data, thereby offering methodological and pedagogical implications for enhancing statistical literacy in diverse classroom settings. Despite these contributions, the study is constrained by the small sample size and the short duration of the instructional intervention. While these limitations restrict generalisability, the evidence nonetheless affirms that TDS-oriented tasks can support durable conceptual growth in students' understanding of central tendency. Future research should adopt longitudinal designs, encompass more diverse learner populations, and extend the focus to additional statistical constructs such as variability and distributional reasoning. From a practical perspective, educators and curriculum developers are encouraged to integrate TDS-informed task sequences that prioritise exploration, argumentation, and reflective decision-making to strengthen students' statistical thinking and promote more robust learning outcomes.

References

Abdüsselam, M. S., Turan-Güntepe, E., & Durukan, Ü. G. (2022). Programming education in the frameworks of reverse engineering and theory of didactical situations. *Education and Information Technologies*, *27*, 6513–6532. https://doi.org/10.1007/s10639-021-10883-8

Barquero, B., & Bosch, M. (2015). Didactic engineering as a research methodology: From fundamental situations to study and research paths. *In Task Design In Mathematics Education: an ICMI study 22, pp. 249-272. Springer International Publishing.* https://doi.org/10.1007/978-3-319-09629-2_8

Bartzia, E.-I., Lodi, M., Sbaraglia, M., Modeste, S., Durand-Guerrier, V., & Martini, S. (2024). An unplugged didactical situation on cryptography between informatics and mathematics. *Informatics in Education*, *23*(1), 25-56. https://hal.science/hal-04184262v1/file/infedu.2024.06.pdf

Ben-Zvi, D., & Makar, K. (2016). *The teaching and learning of statistics: International Perspectives*. Spinger.

Bogas, H., Nascimento, M. M., & Morais, E. (2019). All i have to give – classes to overcome semiotic conflicts in the measures of central tendency in professional education. *In ICERI2019 Proceedings, pp. 141-150. IATED.* https://doi.org/10.21125/iceri.2019.0056

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Kluwer Academic Publishers.

Carlson, B. (2002). Preparing workers for the 21st century: The importance of statistical competencies. *Proceedings of the International Conference on Teaching Statistics (ICOTS-6), pp. 1-6*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9f03ef7ec52d6b4b1e1568083f05bbaa1fd7971e

Conway IV, B., Martin, W. G., Strutchens, M., Kraska, M., & Huang, H. (2019). The statistical reasoning learning environment: A comparison of students' statistical reasoning ability. *Journal of Statistics Education*, *27*(3), 171–187. https://doi.org/10.1080/10691898.2019.1647008

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.)*. Sage Publications.

Dewi, N. R., Yusmaniar, R., Putra, R. W. Y., Alhaq, A., & Suri, F. I. (2024). Analisis learning obstacle materi statistika pada mata kuliah pembelajaran matematika SMP/MTs. *TEACHING: Jurnal Inovasi Keguruan dan Ilmu Pendidikan*, *4*(2), 76-80. https://doi.org/10.51878/teaching.v4i2.2992

Fitri, I., Setyaningrum, W., & Pulungan, D. A. (2023). Fenomena literasi statistik pada pembelajaran matematika siswa SMA di Lhokseumawe Aceh. *Aksioma: Jurnal Program Studi Pendidikan Matematika*, *12*(2), 1927-1941. https://doi.org/10.24127/ajpm.v12i2.7000

Forgasz, H., Hall, J., & Robinson, T. (2024). Evaluating pre-service teachers' statistical literacy capabilities. *Mathematics Education Research Journal*, *36*(1), 231–258. https://doi.org/10.1007/s13394-022-00438-6

Hoffrén, J. (2021). Statistical literacy competencies as post-modern basic civic skills. In *Proceedings of the 63rd ISI World Statistics Congress*, *11*, pp. 489-492.

Hortelano, J. C., & Prudente, M. (2024). Effects of the theory of didactical situations' application in mathematics education: A metasynthesis. *Journal of Pedagogical Research*, *8*(3), 246–262. https://doi.org/10.33902/JPR.202426908

Ismail, Z., & Chan, S. W. (2015). Malaysian students' misconceptions about measures of central tendency: An error analysis. *AIP Conference Proceedings*, *1643*(1), 93–100. https://doi.org/10.1063/1.4907430

Karaca, Ş. A., & Ay, Z. S. (2025). Investigation of eighth grade students' performance on tasks involving statistical thinking about measures of central tendency. *Participatory Educational Research*, *12*(1), 18–42. https://doi.org/10.17275/per.25.2.12.1

Landtblom, K., & Sumpter, L. (2025). Which measure of central tendency is most useful? Grade 6 students' expressed statistical literacy. *Statistics Education Research Journal*, *24*(2), 1-18. https://doi.org/10.52041/serj.v24i2.811

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Lipič, N., & Ovsenik, M. (2020). The effect of statistical literacy on response to environmental change. *Organizacija*, *53*(2), 147-163. https://doi.org/10.2478/orga-2020-0010

Mathé, A.-C., & Mithalal, J. (2025). Figure reproduction as a step towards theoretical geometry: Analysis of the didactical and a-didactical processes in a classroom setting. *Educational Studies in Mathematics*, *119*, 203–223. https://doi.org/10.1007/s10649-024-10370-0

Maryati, I., & Priatna, N. (2018). Analysis of statistical misconception in terms of statistical reasoning. *Journal of Physics: Conference Series*, *1013*(1), 012206. https://doi.org/10.1088/1742-6596/1013/1/012206

Meletiou-Mavrotheris, M., & Lee, C. H. (2024). Revisiting students' statistical reasoning in the age of data science. *Statistics Education Research Journal*, *23*(1), 44–61. https://doi.org/10.52041/serj.v23i1.823

Miles, M. B., Huberman, A. M., & Saldana, J. (2018). *Qualitative data analysis: A methods sourcebook (4th ed.)*. Sage Publications.

Nowell, L. S., Norris, J. M., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, *16*(1), 1–13. https://doi.org/10.1177/1609406917733847

Özmantar, M. F., & Zeynep, U. (2023). Examining the development of students' reflective reasoning within a TDS-based environment. *Journal of Mathematics Teacher Education*, *26*(6), 1127–1149. https://doi.org/10.1007/s10857-022-09542-8

Parra-Fica, J. H., López-Martín, M. del M., & Díaz-Levicoy, D. (2024). Measures of central tendency in primary education textbooks in Chile. *EURASIA Journal of Mathematics*, *20*(6), 1-14. https://doi.org/10.29333/ejmste/14592

Reinhart, A., Evans, C., Luby, A., Orellana, J., Meyer, M., Wieczorek, J., Elliott, P., Burckhardt, P., & Nugent, R. (2022). Think-aloud interviews: A tool for exploring student statistical reasoning. *Journal of Statistics and Data Science Education*, *30*(2), 100–113. https://doi.org/https://doi.org/10.1080/26939169.2022.2063209

Romadoni, G., Sahrudin, A., & Rosdianwinata, E. (2023). Efektivitas pembelajaran menggunakan akun belajar.id Kemendikbud pada materi statistika ditengah pandemi. *Radian Journal: Research and Review in Mathematics Education*, *2*(1), 11-18. https://doi.org/10.35706/rjrrme.v2i1.8708

Safitri, R., & Ruli, R. M. (2025). Analisis hambatan belajar siswa SMK pada materi statistika. *PHI: Jurnal Pendidikan Matematika*, *9*(1), 38-47. http://dx.doi.org/10.33087/phi.v9i1.429

Saidi, S. S., & Siew, N. M. (2018). Assessing students' understanding of the measures of central tendency and attitude towards statistics in rural secondary schools. *Electronic Journal of Mathematics Education*, *14*(1), 73–86. https://doi.org/10.12973/iejme/3968

Schield, M. (2005). Information literacy, statistical literacy, data literacy. *IASSIST Quarterly*, *28*(6), 2–3. https://doi.org/10.29173/iq790

Sulistyani, N., Prasojo, L. D., Purnomo, Y. W., Ardhana, S. S., & Putri, Y. D. (2024). Epistemological didactical situation of learning factors and multiples in elementary school students. *Mathematics Teaching Research Journal*, *16*(5), 66-91. https://files.commons.gc.cuny.edu/wp-content/blogs.dir/34462/files/2024/11/3-Sulistyani-et-al-Epistemological-Didactical-Situation.pdf

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1-8. https://doi.org/10.1080/01621459.1993.10594283

Yenil, T., Arslan, Ç., & Broutin, M. S. T. (2023). Triangle inequality concept teaching: The theory of didactic situations case. *Journal of Pedagogical Research*, *7*(4), 14–29. https://doi.org/10.33902/JPR.202318961

Yin, R. K. (2014). *Case study research: Design and methods (5th ed.)*. Sage Publications.