UNDERGRADUATE STUDENTS' INCONSISTENT ROUTINES WHEN ENGAGING IN STATISTICAL REASONING CONCERNING MODE

DESI RAHMATINA
Universitas Maritim Raja Ali Haji
desirahmatina@umrah.ac.id

NORASYKIN MOHD ZAID
Universiti Teknologi Malaysia
norasykin@utm.my

ABSTRACT

Using the commognitive construct of routine—repetitive rules or patterns observed in statistical discourse—we aimed to investigate how students use inconsistent routines when engaging in statistical reasoning about mode in the context of comparing modes across several data groups. The study data was collected by distributing mode-related questions to students through a Google Form, followed by interviews. Four mode-related questions were given to 43 undergraduate students participating in the study. The results showed that routine plays a significant role in statistical reasoning. The study identified two factors that contributed to the occurrence of inconsistent routines among students: (a) the way students described the data display and (b) the disconnection between routine and endorsed narrative. The results of this research highlight the importance of providing students with opportunities to work with diverse forms and conditions of data associated with mode.

Keywords: *Inconsistent routines; Statistical reasoning; Mode; Data comparison*

1. INTRODUCTION

Routine is closely related to students' methods of reasoning and serves as a significant commognitive component that can be used as a lens in research (Güçler, 2013; Viirman, 2015). Students establish a framework for understanding statistical concepts and solving problems by adhering to routines. These routines help students to develop systematic problem-solving strategies and enhance their understanding of mathematics and statistics. Ioannou (2018) argued that investigating students' use of routine is crucial when examining their difficulties in learning mathematics. Difficulties in learning statistics and probability have been noted by many researchers, such as Konold et al. (1993), who stated that students were inconsistent in their reasoning about probability because they changed their reasons for the way they were solving problems. Garfield et al. (2008) also noted that students' statistical reasoning is often inconsistent from one item or topic to the next, depending on the problem context and students' associated experiences. From a commognitive perspective, Rahmatina et al. (2022) claimed that routine plays an important role in statistical reasoning and might explicate some of the inconsistencies found in student reasoning. As instructors of undergraduate statistics courses, we conjectured that examination of students' use of routines might be a fruitful way of understanding their statistical reasoning processes when engaging with statistical information.

Similar to other introductory statistics methods courses, our students study content to which they have previously been exposed at the high school level. Although teachers consider the mode easier to understand and explain than the median and mean (Landtblom & Sumpter, 2019), research on students' procedural knowledge about the mode is scant compared to similar research about the mean and median. Groth and Bergner (2006) stated that this situation has arisen because there is a perception that the procedure for calculating the mode is less complex than that for calculating the mean and median. They argued that a key area for further research is to identify how to help students develop a deep understanding of the mode. To develop a deep understanding of the mode, learning how to find the mode of one group is not sufficient; rather, comparing modes of multiple data groups is essential for

Statistics Education Research Journal (early access), https://doi.org/10.52041/serj.691

eliciting and developing statistical reasoning about the mode (Biehler et al., 2018; Frischemeier, 2019; Makar & Confrey, 2004; Shin, 2021). Therefore, research is needed to investigate not only the role and nature of routines when students carry out statistical reasoning but also students' reasoning about modes, namely when comparing modes of several data groups. To address this need, the research question for this study was: How inconsistent are the routines that undergraduate students use when performing statistical reasoning about modes?

2. LITERATURE REVIEW

2.1. ROUTINE AS A COMMOGNITIVE COMPONENT

Sfard (2008) introduced commognition as a fusion of communication and cognition aspects. Within a commognitive framework, commognitive components encompass word use, visual mediators, narratives, and routines. Sfard stated that routines in particular serve as regulatory mechanisms for word use, visual mediators, and narratives. She defined routines as a set of meta-rules that describe repetitive discursive patterns in specific situations, representing recurring outlines inherent in a given discourse. That is, routines are patterned ways of doing things, such as defining or proving (Sfard, 2020). Since Sfard (2008) introduced commognition as a way of understanding students' reasoning patterns, many other researchers have added their descriptions and insights into how routines are manifested in learning situations. Lavie and Sfard (2019) explained that routines encompass ways of accomplishing tasks that are characterized by regular and explicit habits exhibited in a specific community, such as defining, hypothesizing, proving, estimating, generalizing, and abstracting (Nardi et al., 2014). Routines become evident when students provide detailed explanations of problem-solving steps, document existing knowledge, break down problems into manageable parts, and identify set goals (Zayyadi et al., 2019).

The relationship between routines and narratives is mentioned by Kotsopoulos et al. (2009) and Mpofu and Pournara (2018). Kotsopoulos et al. (2009) recognized the relationship between routines and endorsed narratives, indicating that routines are used to generate narratives endorsed by individuals. Endorsed narratives are sequences of spoken or written utterances framed as a description of objects, relations between objects, or activities with or by objects (Tabach & Nachlieli, 2016). An example of an endorsed narrative is a definition or a theorem. Mpofu and Pournara (2018) described the reasons students give for the ways they solve problems as substantiation narratives. For example, when students choose the largest mode when comparing several groups of data, they can substantiate their choice by giving the reason that its value has the highest frequency. Disruptions in the relationship between routines and narratives can lead to errors, which can result in students experiencing cognitive conflicts (Pratiwi et al., 2022).

Furthermore, routines are perceived as the procedures or steps learners adopt to solve problems, which emerge as the outcome of activities aimed at producing something, with narratives serving as a tool in this process. Lavie and Sfard (2019) categorized routines into two types: (a) practical and discursive routines and (b) process-oriented and product-oriented routines. A routine is practical if someone interprets a task as something that requires a change or repositioning of objects. Daily routines might include preparing breakfast. When preparing breakfast, one might consider what breakfast needs to be prepared and decide whether to repeat the choice made the previous day or whether a change is needed. A routine is discursive if someone interprets a task as requiring communication. Lavie and Sfard (2019) defined a discursive routine as a pattern followed when communicating with others or with oneself. For example, one might communicate with oneself to select a method to complete a task. Communication with oneself is referred to as thinking (Lavie & Sfard, 2019; Sfard, 2018). Ng (2016) gave an example of a routine in a learning situation, such as when a teacher uses certain words or gestures repeatedly to model a discursive pattern. A process-oriented routine focuses on performance, not results. Conversely, a result-oriented routine focuses on the results of an activity where new objects are created or rearranged. Lavie and Sfard (2019) also revealed that whatever we do involves routines and activities carried out from the simplest to the most sophisticated, such as repetition of something that has been done or seen before to something exploratory that transforms a ritualized routine into an exploration producing a new endorsed narrative and shifting the attention from processes to products. A repetition also might differ from what has been done before, which is referred to as an inconsistent routine. The inconsistencies associated with creating steps required to solve problems are closely related to students' reasoning.

Studying students' inconsistent uses of routines has proven useful in understanding their learning and reasoning processes in many disciplines. In the context of calculus instruction, Güçler (2013) focused on analyzing routines—conceived as patterns governed by meta-level rules—in the discourse of a lecturer and undergraduate students for the concept of limits. She found that students' difficulties often occurred in contexts where the instructor shifted elements of their discourse (such as word use, metaphors, or representations) without making those shifts explicit, which led to communication breakdowns and misunderstandings. Viirman (2015) examined learning routines when college lecturers taught functions. His findings showed that different types of explanation routines emerging in the discourses of teachers are due to a tendency towards an inverse relationship between the use of mathematical facts and everyday language. According to Ioannou (2018), misapplication of meta-rules in mathematical routines can result in commognitive conflicts. Such meta-level difficulties were evident in his research, even at the early stages of a teaching module, where students' object-level understanding was found to be better than their understanding and application of required meta-rules. Other researchers, such as Tabach and Nachlieli (2015), have used routines to demonstrate changes in students' discourse about functions, whereas Fernández-León et al. (2021) have identified routines in the discourse of undergraduate students when describing and defining solids. However, the utility of routine extends beyond the learning process. Routine has also been used in research related to mathematical reasoning (Jeannotte & Kieran, 2017), geometric reasoning (Toscano et al., 2019; Wang & Kinzel, 2014), and statistical reasoning (Park & Lee, 2014; Rahmatina et al., 2022).

2.2 STATISTICAL REASONING ABOUT MODE

Statistical reasoning is crucial to understanding and interpreting daily phenomena (Bennett et al., 2017). According to Bargagliotti et al. (2020), the statistical problem-solving process is the foundation and core of statistical reasoning and making sense of data. They describe the statistical problem-solving process as consisting of four components: (a) formulating a statistical investigative question, (b) collecting or considering data, (c) analyzing data, and (d) interpreting results. Within this problem-solving process, statistical reasoning operates across all four components because it often involves sustained attention to variability, context, and the purpose of the investigation. When formulating a statistical investigative question, students are expected to consider which aspects of the data may be meaningful, what variability might be anticipated, and how the question could be reasonably answered using the data. Similarly, while collecting or considering data, statistical reasoning may be reflected in how students evaluate the quality of the data, how the data were obtained, and the extent to which the data appear suitable for addressing the investigative question. In practice, statistical reasoning entails making interpretations based on data, data representations, or statistical summaries of data (Garfield & Chance, 2000; Ben-Zvi, 2004). It also involves interpreting and drawing meaningful conclusions from datasets (Martin et al., 2009). Specifically, Jones et al. (2004) stated that students' statistical reasoning refers to four statistical processes, namely describing data, organizing data, representing data, and analyzing and interpreting data. In this study, we focus on describing data. This is because a student's ability to read data displays is fundamental to making predictions and finding trends (Jones et al., 2004). Furthermore, how one makes predictions is influenced by how one reads a data display.

Another foundational aspect of statistical reasoning is conceptual understanding of key ideas such as distribution, center, and spread (Garfield, 2002). To develop conceptual understanding, students need to have a deep understanding of the foundational elements of concepts before making connections among elements or among concepts. Students' overarching conceptual understanding can be measured through their ability to recognize and explain important concepts in a domain (Rittle-Johnson & Star, 2009). In the context of the mode, students first need to understand the elements involved, including the distinction between the concept of a variable and the concept of frequency. Knowing the definition of the mode—the value that appears with the highest frequency in a set of data (Mann, 2013)—is insufficient for developing connection knowledge. Crooks and Alibali (2014) identified six types of definitions of conceptual understanding. The most common is connection knowledge, which refers to understanding relationships and connections within a domain, such as relationships among ideas (e.g., Hiebert & Lefevre, 1986). To apply this connection knowledge, students need to relate the concept of mode to the data represented, ensuring that their reasoning aligns with the definition of mode rather than relying on surface features of data representations when comparing modes across several data groups. Reasoning with the definition of mode in relation to the representation is crucial for correctly identifying the group with the largest, smallest, or equal mode among several data groups.

Several studies have examined how learners understand and use measures of central tendency. For example, Groth and Bergner's (2006) study used the Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs & Collis, 1982) as a lens to identify the level of thinking of preservice elementary and middle school teachers when comparing mean, median, and mode. Their findings revealed that most preservice elementary and middle school teachers exhibited unistructural and multistructural levels of thinking. In the SOLO taxonomy, the unistructural level refers to responses based on only one relevant aspect, such as explaining that the mode is the value that appears most frequently in a dataset, where students consider only the frequency of occurrence without taking the numerical value into account. Unistructural responses result in a limited and potentially dogmatic conclusion. In contrast, the multistructural level involves identifying several relevant aspects, such as considering both the values and their frequencies when determining the mode. However, these aspects have not yet been integrated into a coherent understanding of how to compare the modes across multiple data groups. These multiple aspects are considered independently and are not yet integrated, which may lead to firm yet superficial conclusions (Biggs & Collis, 1982). Groth and Bergner argued that a key area for further research was investigating how to help preservice teachers develop a deep understanding of the concepts of mean, median, and mode. There is limited research on students' understanding of the mode, possibly because it is perceived to be an easy concept to identify. Indeed, the first author, a lecturer of a statistical methods course, asked her undergraduate students to identify which concept was the easiest to grasp among average, median, and mode. Of the 43 students, 90% stated that the mode was the easiest concept to grasp. They reported that identifying the mode simply involved determining the value that occurred most frequently. In a question given to them, they were asked to identify the mode of the math scores of 10 students, which were 60, 60, 65, 70, 75, 75, 75, 80, 80, and 85. All students correctly identified the mode as 75 because the majority were awarded a score of 75 compared to the others. Thus, our initial studies found that students considered the concept of mode easy to grasp and calculate when faced with a single data group.

Leavy and O'Loughlin (2006) investigated prospective elementary teachers' conceptions of the mean and examined how they understood and used it as a statistical measure. In their study, 57% of the prospective teachers used the mean correctly to compare two datasets. Moreover, Karatoprak et al. (2015) found that the prospective elementary and secondary school mathematics teachers in their study chose the mode as the average without considering whether the variables were categorical or not, and without considering intervals, ratios, or outliers when examining the data. The use of the average was determined according to the type of data and the presence of outliers. Karatoprak et al. (2015) reported that prospective teachers mistakenly chose the mode rather than the mean as the appropriate measure of central tendency to represent a dataset due to limited understanding of various types of variables. This challenge is not limited to prospective teachers. Bakker (2003) found that although students were familiar with the term 'average' in its everyday use as 'typical,' they had not yet developed a technical understanding of it as a representative value. Their nontechnical understandings might stem from limited awareness of the formal properties of the mode. Landtblom (2023) highlighted that the mode has two unique mathematical characteristics: first, a dataset can have zero or multiple modes; and second, the mode is the only measure suitable for nominal data. Hence, to be able to reason well when comparing modes, sufficient knowledge is needed about related concepts, such as no mode, unimodal, bimodal, multimodal, and the type of data. We conjecture that comparing the modes of multiple data groups may involve higher-level reasoning than identifying the mode of a single group.

The procedure, method, or routine used to compare the modes of several data groups, especially when data is displayed in various forms such as summary tables, graphs, and ordered data points, should be a primary concern for students. For example, students can determine the largest mode value from several groups of data by 1) determining the mode of each group, 2) comparing the mode of one group with that of another group, and 3) determining the largest mode from several groups of data based on the results of the comparison. Students can use a similar procedure to determine the smallest mode of several data groups. In this research, the term "routine" specifically refers to the procedure applied to compare the modes of different data groups. This study aimed to investigate undergraduate students' inconsistent routines when engaging in statistical reasoning about modes to compare multiple data groups. The manifestation of statistical reasoning was observed through the meaningful comparison of modes in different data groups. Statistical reasoning was viewed in terms of how undergraduate students described the data when comparing modes in different groups of datasets. Inconsistent routines resulted from undergraduate students *changing* the steps taken to compare the modes displayed in various types of datasets.

The data in this study were presented in the form of graphs, tables, and ordered data points. van Garderen et al. (2014) observed that "reasoning with diagrams is a challenging process; hence, students need more time and experience to develop [this skill]" (p. 13). Accordingly, the authors assumed that the steps taken by undergraduate students when comparing modes in different data groups reflected their reasoning as they interpreted the displayed data. Consequently, students who face difficulties in reasoning are likely to find it difficult to perform explanation tasks when comparing modes, leading to inconsistency in the steps used to compare modes in various data groups. Comparing groups motivates the need for and use of graphical displays, which are best employed for showing differences between groups but are not easily understood or interpreted by students (Garfield et al., 2008). In this study, the graphical displays presented to students serve as visual mediators to facilitate students' reasoning when comparing mode values across data groups. Visual mediators, such as graphs, mediate ideas and often influence what one can say about the idea discussed (Tabach & Nachlieli, 2011). To interpret graphical displays accurately, students need to understand how to read graphs. Friel et al. (2001) noted that "to read the information directly from a graph, one must understand the conventions of graph design, and to manipulate the information read from the graph, one makes comparisons and performs computations" (p.152). Despite this potential, students may encounter difficulties in interpreting the displays accurately.

3. RESEARCH METHODOLOGY

3.1. INSTRUMENTATION

Four assessment items focused on modes (see Figure 1) were used to explore the inconsistent routines used by undergraduate students when engaging in statistical reasoning to compare modes of several groups of data. Data was displayed in the form of graphs, summary tables, and ordered data points. These items were presented in a multiple-choice format, and students were then asked to briefly elucidate the reasoning underlying their selected answer. Each multiple-choice item included five to six response options, with the last two options being *the mode cannot be determined from the given information* and *I do not know.* The inclusion of the *I do not know* option allowed students to indicate their inability to provide an answer. The four items were developed based on materials about data presentation in the form of frequency tables, graphs, or diagrams, which were taught to undergraduate students before modes were introduced. In the first item, quantitative data was presented in a summary table, which required the undergraduate students to select the group (out of three groups) with the largest mode. In the second item, quantitative data was presented as dot plots, which required undergraduate students to select the group with the same mode size. In the third item, quantitative data was presented as ordered data points, and students were asked to choose which group (out of three) had the largest mode. In the fourth item, quantitative data was presented as ordered data points, and students had to select which group out of the two groups had the larger mode. Interview guidelines were also used to gain deeper insights into the routines performed by students.

Item Number

1

Class A

Score	Frequency
60	5
70	10
80	10
90	4
100	1

Class B

Score	Frequency
60	2
70	6
80	9
90	10
100	3

Class C

Score	Frequency
60	6
70	6
80	6
90	6
100	6

Identify the class with the largest mode in mathematics scores awarded to the students.

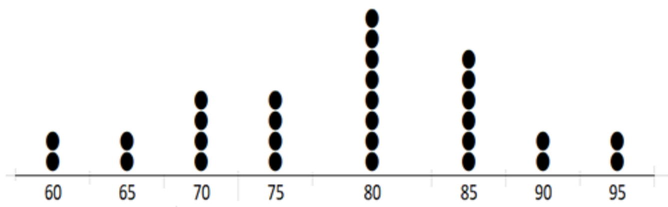
a. No class has the largest mode.
b. Class A has the largest mode among the other classes.
c. Class B has the largest mode among the other classes.
d. Class C has the largest mode among the other classes.
e. The mode cannot be determined from the given information.
f. I do not know.

Give reasons for your answer.

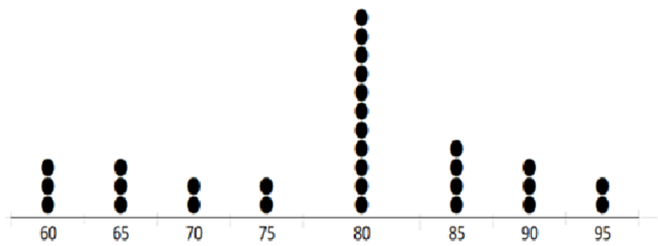
2

The following graph shows the distribution of statistics scores for students in two classes.

Class A



Class B



The conclusion about the mode deduced from the two graphs is:

a. The statistics scores of Class A have a larger mode than Class B.
b. The statistics scores of Class B have a larger mode than Class A.
c. The statistics scores of Classes A and B have the same mode.
d. The mode cannot be determined from the given information.
e. I do not know.

Give reasons for your answer.

3

The following data shows the salaries (in hundreds of thousands) of 15 employees in three companies.

Company A	35	35	37	37	40	40	40	40
	42	42	43	43	45	45	50	
Company B	35	35	37	37	40	40	41	42
	42	45	46	50	50	50	50	
Company C	35	35	35	35	40	40	41	42
	42	45	46	48	48	50	50	

The conclusion about the mode from the aforementioned data is:

a. No company has the largest mode.
b. The salaries of employees in Company A have the largest mode among the other ones.
c. The salaries of employees in Company B have the largest mode among the other ones.
d. The salaries of employees in Company C have the largest mode among the other ones.
e. The mode cannot be determined from the given information.
f. I do not know.

Give reasons for your answer.

4

The distribution of mathematics scores for 15 students in two classes is stated as follows.

Class A	60	60	60	60	60	65	65	70
	70	70	75	75	80	85	90	
Class B	60	60	65	70	70	75	75	78
	80	85	85	85	85	90	90	

The conclusion about the mode of mathematics scores awarded to the students in the two classes is stated as follows:

a. The mathematics scores in Class A have a larger mode than B.
b. The mathematics scores in Class B have a larger mode than A.
c. The mathematics scores in Classes A and B have the same mode.
d. The mode cannot be determined from the given information.
e. I do not know.

Give reasons for your answer.

Figure 1. Items about the mode given to students (correct response is italicized)

3.2. PARTICIPANTS

The participants consisted of 43 undergraduate students, 11 males and 32 females, between the ages of 18 and 22 years. These undergraduate students were enrolled in a second-semester statistics course in the mathematics education program at a state university in Tanjungpinang, Indonesia, and all were prospective secondary teachers. All participants were given a test that required them to engage in statistical reasoning in comparing modes in several data groups. After students completed the test, we interviewed them to gather information about the routines they followed. Their routines were categorized into three types: (a) consistent routines using the correct approach, (b) consistent routines using an incorrect approach, and (c) inconsistent routines. Inconsistent routines occurred when undergraduate students correctly followed the steps for comparing modes in one item but made mistakes in others. Based on the test results and interviews, three undergraduate students who demonstrated inconsistent procedures were selected as research subjects. These students compared modes using the correct procedure for one, two, or three items but applied an incorrect procedure for the remaining items. The reason for not selecting students who were correct or incorrect in comparing modes for the four items is that these students were likely to use consistent routines in comparing modes for all items. Thus, the aim was to select students who exhibited inconsistent routines when comparing the modes of multiple data groups. Of the four items given, Student 1 used the correct routine for three items, Student 2 used the correct routine for two items, and Student 3 used the correct routine for one item. The student retrieval in this study is depicted in Figure 2.

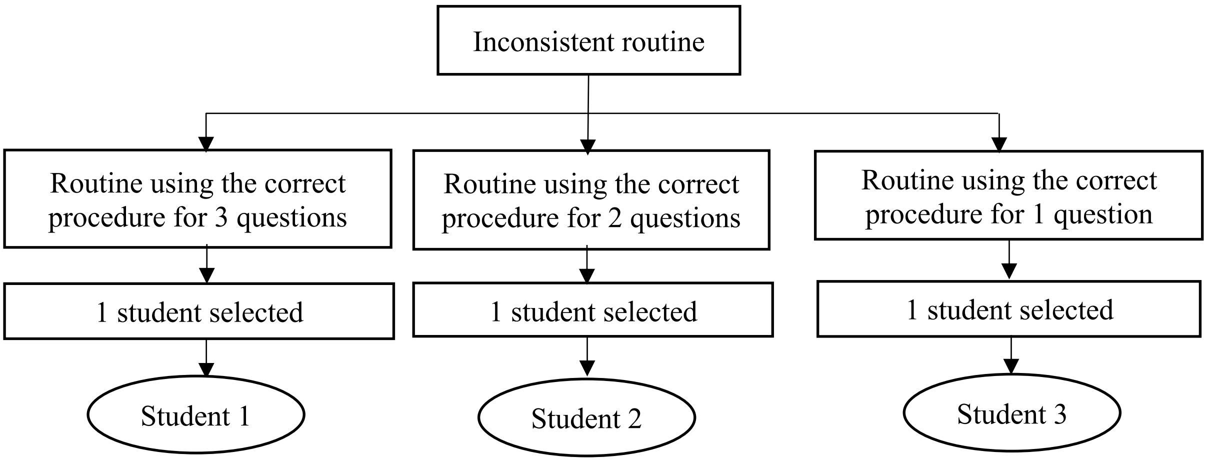


Figure 2. Selection of students in this study

3.3 DATA COLLECTION

Data collection involved obtaining information through students' responses to mode-related items and through interviews with the participants. The students responded to the questions after receiving instruction on measures of central tendency, including mean, median, and mode. Before studying the measures of central tendency, students had also received instruction on data representation using various graphs such as histograms, bar graphs, stem-and-leaf plots, scatter diagrams, and dot plots. The mode-related items were administered using Google Forms. The first author, who also served as the lecturer for the statistics methods course, facilitated the administration of the mode-related items. After the students completed the items, interviews were conducted individually with each of the three participants to gather more comprehensive insights into their inconsistent routines. A semi-structured interview guide, which included items designed to elicit specific answers, was used to enable us to explore the inconsistent routines students used when performing statistical reasoning. The interview guidelines were adjusted to the conditions at the time of the interview, which included questions beyond those that had been designed. Magaldi and Berler (2020) explained that a semi-structured interview approach can generate a great deal of information that can also be complex. During the interviews, the students were asked (a) which procedure they used to determine whether one group of data was larger than another group; (b) why they chose a particular answer out of the options provided; (c) what the mode was of each group; and (d) what a mode was.

Data triangulation, a method of validating, challenging, or expanding upon existing findings, was performed to ensure data validity (Turner & Turner, 2009). In this research, triangulation was achieved by checking the consistency between test answers and interview transcripts. The first author, a statistics lecturer, conducted interviews with the undergraduate students after they completed the items related to mode.

3.4 DATA ANALYSIS

The data analysis involved four main stages. The first stage entailed data coding, which was performed by the first author. Data coding was carried out based on students' answers to four items that compared the modes of several data groups. There were four data codes, namely (a) student chose the correct option and provided a rationale corresponding to the mode concept, (b) student chose the correct option but could not provide a rationale based on the mode concept, (c) student chose an incorrect option and could not provide a rationale based on the mode concept, and (d) student work did not fall into any of the previous codes, such as if a student chose the "I do not know option" and did not give reasons. In the second stage, the students' responses were grouped based on data coding. In the third stage, the students' responses were grouped into consistent and inconsistent routines based on interview results. Lastly, the inconsistent routines exhibited by students when comparing modes across multiple datasets were described. Inconsistent routines meant that students correctly followed the steps for comparing modes in one item but took incorrect steps in other items. For example, the procedure or steps students used to compare modes in item 1 were correct, but they were incorrect for other items.

4. RESULTS

A total of 43 students were administered the items regarding the comparison of modes across multiple datasets. Out of the 43 students, one student followed consistent routines correctly, 13 students followed consistent routines incorrectly, and 29 students followed inconsistent routines. Three out of 29 undergraduate students were selected as research subjects, with pseudonyms of Rizi, Almi, and Hafa. The three undergraduate students carried out inconsistent routines, using statistical reasoning to compare the modes of several data groups. The reasons given by students for choosing one of the options are referred to as substantiation narratives.

4.1. RIZI

Rizi chose the correct option for items 1, 3, and 4. However, for item 2, Rizi chose the wrong option. The reasons given by Rizi for choosing the given answers to items 1 to 4 are as follows.

Item 1: Based on the data provided, it was concluded that Class B has the largest mode value of 90. *The mode represents the value that occurs most frequently in a dataset.* Upon examining the table for the three classes, *it is evident that 90 appeared 10 times in Class B,* indicating that it is the most common value among the options.

Item 2 : In accordance with the diagram, it is evident that Class B showed the highest mode and *represents the most frequently occurring data point*. Specifically, *the mode value is 80, with a frequency of 11 instances.*

Item 3 : Company B has a *frequency of 4, which means it occurred four times*, specifically at a *value of 50*.

Item 4 : The mode in Classes A and B is *60 and 85*, respectively. Therefore, *class B has a mode greater than A.*

Rizi chose option C for item 1 and option B for item 2. Hence, Rizi correctly provided an endorsed narrative, defining mode as "the most frequently occurring data point." However, the statistical reasoning utilized by Rizi when using the data display to compare modes differed between the data presented in summary table form in item 1 and the data presented in graph form in item 2. Furthermore, the substantiation narrative provided by Rizi for choosing option C for item 3 and option B for item 4 shows that the statistical reasoning used to describe the data presented in the form of ordered data points is the same in both items 3 and 4. Hence, it is shown that Rizi was inconsistent in using a substantiation narrative when choosing one of several data groups.

In items 1, 3, and 4, Rizi was able to provide correct reasoning to compare modes. However, in item 2, Rizi could not give the same reasoning as for the other items. For items 1, 3, and 4, Rizi focused on the values of the mode for comparison. In contrast, in item 2, his reasoning shifted towards a frequency comparison rather than comparing mode values.

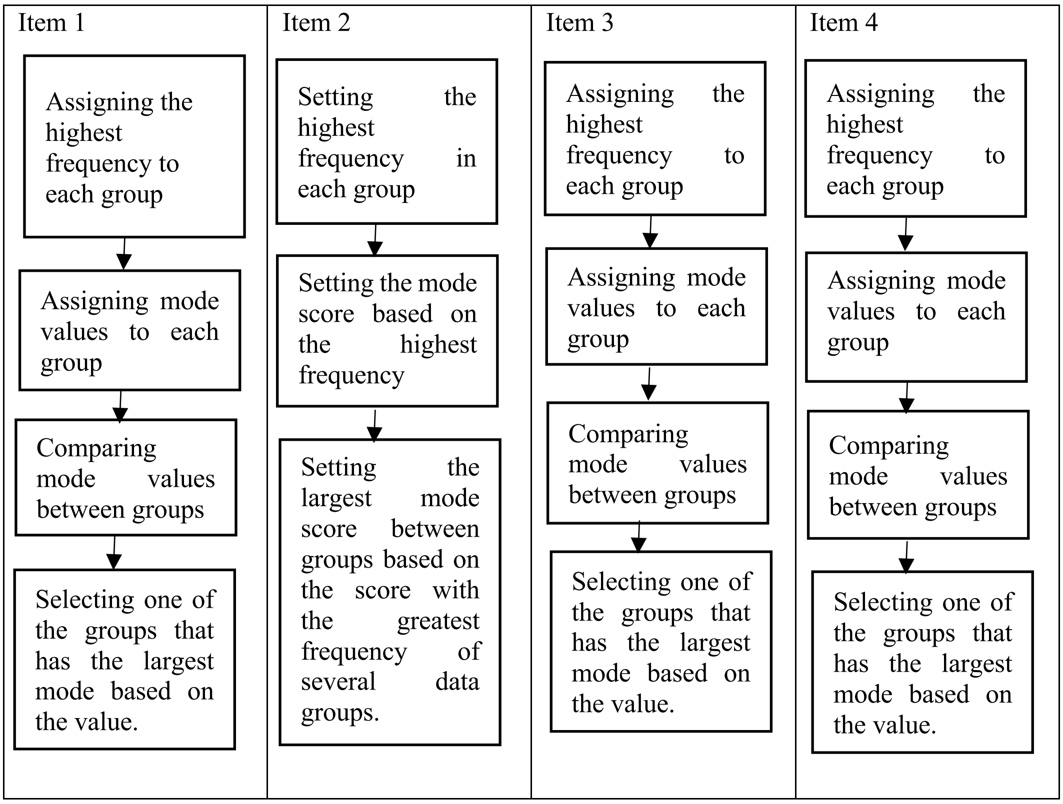


Figure 3. Routines performed by Rizi when engaging in statistical reasoning about the mode

Figure 3 shows the mode comparison process adopted by Rizi in response to the four items. Rizi followed a consistent routine in problems 1, 3, and 4, where he initially compared mode values across multiple datasets and then determined the largest one. In contrast, in item 2, where the data were presented as a graph, his procedure differed—he focused mainly on frequency rather than on the mode value itself. This indicates that Rizi applied different routines depending on the data display format, showing procedural variation. This observation was supported by the interview between the researcher (R) and Rizi.

R: How did you determine that class B has a larger mode value than A in question 2?
Rizi: Initially, I *examined the dot plot diagram* to visualize the data distribution in both classes. … Then, *I compared the number of data points for each value within the classes*. By identifying the mode value in each class, I concluded that class B had the highest mode value of 80 with a count of 11 *data points*.
R: What is the mode value for class A?
Rizi: The mode value for class A is *80*, and it appeared *8 times*.
R: Can it be considered a mode?
Rizi: Yes, *a mode is defined as the value that appears most frequently in a given dataset*; therefore, in this case, 80 is indeed the mode for class A.

According to the interview results, Rizi recognized a similarity between the mode values of classes A and B in item 2. However, Rizi identified Class B as having a larger mode than Class A. This decision was influenced by the number of data points on the graph representing Class B (11) compared to the number of data points on the graph representing Class A (8) (see Figure 1).

The written answers showed that the substantiation narrative given by Rizi varied when choosing different options in items 1 to 4. The interview results also showed that Rizi followed different procedures or steps in comparing the modes of several data groups. Rizi could define mode correctly, but he could not consistently use the definition of a mode to compare the modes of several data groups. This outcome shows that the varying routines in comparing modes was due to the substantiation narrative expressed by Rizi about different modes when choosing among several data groups. Furthermore, in one data display condition, Rizi had correctly connected his mode definition with how to describe the data display. Rizi compared the mode by first determining the mode in each data group and then comparing the mode values as the basis for comparing the modes of several data groups. However, in other data display conditions, Rizi compared the modes of several groups of data. He did not compare the values but rather the frequency of the mode values obtained in each group of data. This shows that Rizi used a different procedure when comparing the modes but retained the same definition of the modes. A recapitulation of the inconsistent routine employed by Rizi is shown in Table 1.

Table 1. Routine recapitulation performed by Rizi

Item Number	1	2	3	4
1	-	Inconsistent	Consistent	Consistent
2	Inconsistent	-	Inconsistent	Inconsistent
3	Consistent	Inconsistent	-	Consistent
4	Consistent	Inconsistent	Consistent	-

As shown in Figure 3 and Table 1, Rizi used the same procedure to compare the modes in items 1, 3, and 4. In item 2, however, Rizi did not repeat the same procedure used to compare the modes in the previous procedure (item 1). Meanwhile, to compare the modes in items 3 and 4, Rizi repeated the same procedure as used in the previous procedure (item 1). This indicates that Rizi applied inconsistent routines when comparing the modes of several data sets. Although three items were answered using appropriate procedures, different procedures were used in the other item. This reflects a lack of procedural consistency in Rizi's statistical reasoning.

4.2. ALMI

For items 1 and 3, Almi chose the correct response; however, for items 2 and 4, Almi chose the incorrect response. The reasons given by Almi for choosing the various options in items 1 to 4 are as follows.

Item 1 : This is because class B has a *mode of 90*, with as *many as 10 people*, while class A has a mode of *80 and 70***.** Class C has *no mode* because each value has an equal frequency. Based on this data, it was concluded that *class B possesses a mode value greater than the other classes.*

Item 2 : The reason for this is that, according to the provided diagram, class B has a greater mode compared to A. Specifically, the *mode of class B is 80*, with a *count of 11 individuals.*

Item 3 : Company B has the highest mode among the given companies, as it has a *value of 50,* with *4 employees*. In comparison, company A has a *mode of 45, with 4 employees,* while *C has 35, with 4 employees.* Therefore, the mode in *company B is the largest.*

Item 4 : This is because class A has a mode *value of 60,* with a frequency of *5 individuals.*

The substantiation narrative provided by Almi for choosing responses in the four items shows that she correctly used reasons consistent with the endorsed narrative about the concept of mode in items 1 and 3, but not in items 2 and 4. In the correct responses, she provided reasons aligned with the accepted definition of mode as the value that occurs most frequently in a dataset. However, in items 2 and 4, her substantiation narrative differed from the endorsed narrative, suggesting that her justification was influenced more by the form of data display than by the concept of mode itself. In items 1 and 3, Almi's statistical reasoning was accurate when she described the data displays by comparing mode values across datasets, but it differed in items 2 and 4 when she focused mainly on frequency rather than on the mode values themselves. This indicates that her reasoning varied depending on the data display format. For instance, in item 1, she identified Class B as having the largest mode (90) by comparing the values in the summary table, whereas in item 2 she concluded that Class B had the largest mode because 11 students scored 80 in the dot plot. This shift shows that her reasoning was seemingly influenced more by perceptual cues in the graph than by the actual mode values. In items 1 and 3, Almi consistently followed a similar routine to compare the modes across data groups. She identified the mode in each group and compared their values to determine the largest mode. In contrast, in items 2 and 4, her procedure differed—she focused mainly on frequency rather than on the mode value itself. This shows that Almi applied different routines depending on the data display format, indicating procedural variation. This observation is supported by the excerpt from the interview transcript between the researcher (R) and Almi.

R: Why did you conclude that class B has the largest mode compared to the other classes in question 1?

Almi: From the *table* given, it was observed that each class has a different mode. In class A, there are 2 modes, namely 80 and 70, while in B, the mode obtained is 90, and none in C. Therefore, the largest mode is 90 in class B.

R: In question 2, why did you choose class B as having the largest mode compared to the other classes?

Almi: By looking at the data or *graph* given in the question. From the *graph,* it is evident that each class has 8 values, namely 60, 65, 70, 75, 80, 85, 90, and 95. In the *graph* for *class A,* the mode of the statistics values is 80, with eight students attaining that score. Similarly, the *graph* for *class B* also reveals that the most frequent statistics value obtained by students is 80, with *11 individuals* achieving this score. Therefore, based on the given *graphs*, it was concluded that *class B has the mode* of 80, attained by 11 students.

R: What is a mode?

Almi: *A mode refers to the value that occurs most frequently*, is obtained by the most number of individuals, or appears with the highest frequency.

Based on the interview results, it is apparent that Almi was able to define mode correctly as "the value that occurs most frequently," yet Almi used different routines for determining the largest mode in items 1 and 2. In item 1, Almi relied on the mode value to identify the largest mode. However, in item 2, Almi considered the mode values in both classes but ultimately based the determination of the largest mode on the mode with the highest frequency in each class. It is important to note that the decision-making process employed by Almi in item 2 was influenced by the shape of the graph, leading to the selection of the class with the highest frequency (i.e., the class with the most data points on the graph) as the one with the largest mode value. Almi exhibited different routines in choosing the group with the largest mode, as shown in Figure 4.

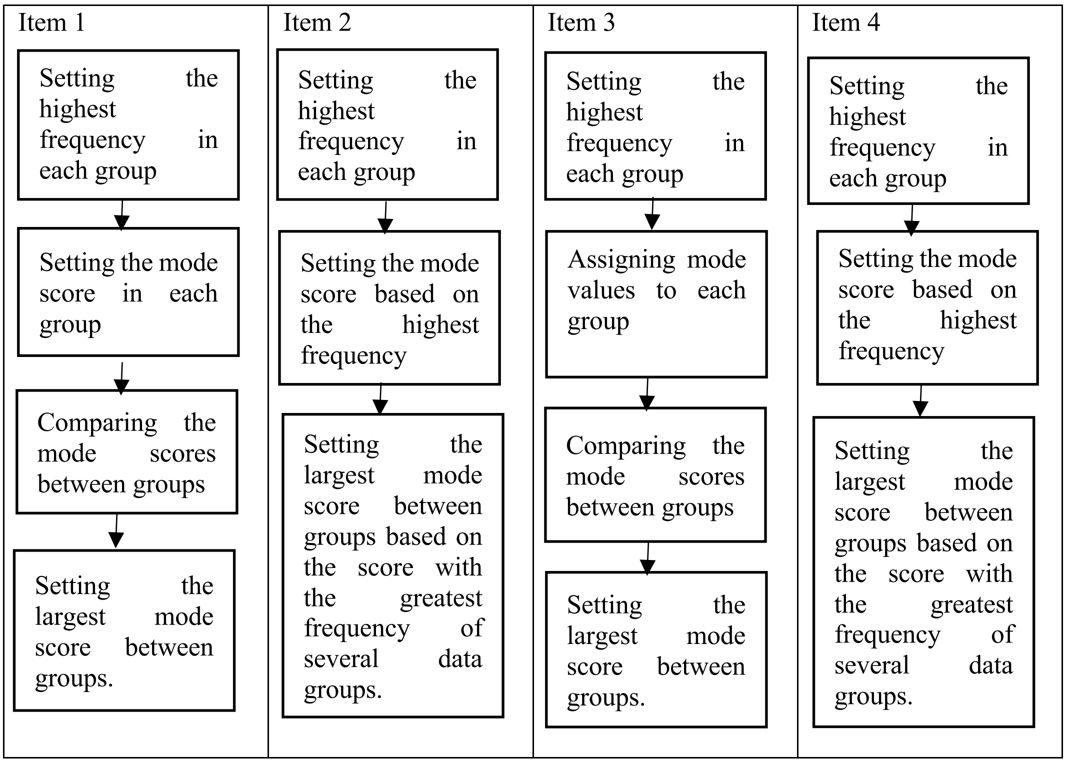


Figure 4. The routines performed by Almi when engaging in statistical reasoning about the mode

The inconsistent routine employed by Almi was observed in the comparisons between items 1 and 2; 1 and 4; 2 and 3; and 3 and 4. A summary of routines performed by Almi is shown in Table 2, detailing the different approaches used in each item.

Table 2. Routine recapitulation performed by Almi

Item Number	1	2	3	4
1		inconsistent	consistent	inconsistent
2	inconsistent		inconsistent	consistent
3	consistent	inconsistent		inconsistent
4	inconsistent	consistent	inconsistent	

As shown in Figure 4 and Table 2, Almi used the same procedure when comparing modes in items 1 and 3, as well as in items 2 and 4. The procedure Almi used was correct for comparing modes in items 1 and 3. However, it was different for items 2 and 4. Almi could not use the correct procedure to compare the modes of several groups of data. Although Almi's procedure was correct for items 1 and 3, she did not repeat the procedure to compare the modes in items 2 and 4. This indicates that Almi applied inconsistent routines when comparing the modes of several data sets. Although two items were answered using appropriate procedures, she used different procedures for the remaining items. This reflects a lack of procedural consistency in Almi's statistical reasoning.

4.3. HAFA

Hafa chose the wrong option for items 1, 2, and 4. However, for item 3, Hafa chose the correct option. The reasons expressed by Hafa for choosing the answers to items 1 to 4 are as follows:

Item 1 :	Class A has the *highest frequency*, which is 10. It is derived from two distinct scores, specifically *70 and 80*. On the other hand, class B also has a *frequency of 10* and is represented by *a score of 90*.
Item 2 :	The rationale behind choosing *class B as having the largest mode* is rooted in the notable dissimilarity observed in the diagram. Specifically, *a score of 80* appeared *11 times,* whereas, in class A, the same score appeared *only 8 times*. It was deduced that class *B possesses a higher mode than A*.
Item 3 :	Considering the given data, it was observed that *Company A has a mode score of 40,* with a *frequency of 4*. Meanwhile, *company B has a mode score of 50* with a *frequency of 4*, while *C has a mode score of 35*, with a *frequency of 4.* Given that *Company B possesses the mode score of 50,* which is the highest, it was chosen.
Item 4 :	Examining the gathered data, it is evident that classes A and B have mode scores of *60 and 85 with frequency of 5 and 3.* As a result, it was concluded that *class A possesses a larger mode compared to B.*

The substantiation narrative expressed by Hafa in selecting option B in items 1 and 2, and option A in item 4, indicates that her statistical reasoning remained consistent across different data representations: summary tables (item 1), dot plots (item 2), and ordered data points (item 4). In all three cases, Hafa described the data representations to compare modes by focusing on frequency across groups. In item 1, Hafa initially described the frequencies of each class, noting that both Class A and Class B had the same maximum frequency (10). She was then seemingly influenced by the fact that Class A had two modes (70 and 80), and both occurred with this maximum frequency. In this case, Hafa chose the class that had more than one mode, as Class A was bimodal. She did not consider the higher mode value of 90 in Class B when selecting the class with the largest mode**.** In item 2, Class B had a higher frequency than Class A and was therefore selected as having the largest mode. In item 4, Class A had a higher frequency than Class B and was selected accordingly. These patterns suggest that Hafa focused on frequency rather than mode values when comparing data groups.

However, although the data representations in items 3 and 4 were both in the form of ordered data points, the substantiation narrative expressed by Hafa reveals a different line of statistical reasoning for item 3 from that in item 4. In item 3, Hafa described the data representation to compare the mode values of each group. Company B had the highest mode value (50) among the companies (Company A = 40, Company C = 35) and was therefore selected as having the largest mode. This indicates that the substantiation narrative constructed by Hafa in selecting option C in item 3 reflects an appropriate use of the endorsed narrative for mode comparison and a correct application of the concept of mode in selecting among several data groups. This inconsistency in reasoning suggests that Hafa did not consistently apply the same substantiation narrative when choosing between multiple data groups.

In item 1, Hafa determined that Class A had the largest mode because Class A had the highest frequency for multiple values compared to the others. Classes A, B, and C had two modes (70 and 80), one mode (90), and none, respectively. Therefore, Class A was selected as having the largest mode (bimodal). In this case**,** Hafa seemed to compare the modes based on the number of modes, rather than the mode values. Although both Class A and Class B had the same maximum frequency (10), she identified Class A as having the highest mode, even though its mode values (70 and 80) were lower than the mode value of 90 found in Class B. In item 2, Hafa chose class B as having the largest mode due to the higher frequency of its score compared to A. Despite Classes A and B having a mode of 80, Class B had a higher frequency (11 students) than A (8 students). Class B has the larger mode based on frequency comparison. In item 3, Hafa determined that Company B had the largest mode among the three companies by apparently considering the highest mode score of employee salaries. The modes of employee salaries in Companies A, B, and C were 40, 50, and 35, respectively. Company B was selected as having the largest mode compared to the other groups. In this case, Hafa seemed to first compare the mode within each group before determining the largest mode among several data groups. Lastly, in item 4, Hafa provided similar reasoning to items 1 and 2, selecting Class A as having the largest mode, likely due to its higher frequency than Class B. The modes in Classes A and B were 60 and 85, respectively, although Class A (5 students) had a larger frequency than Class B (4 students). This led to the conclusion that Class A had the larger mode. The excerpt of the interview transcript between the researcher (R) and Hafa is reported as follows.

R: How can one determine that class B has a larger mode than class A in question 2?
Hafa: It is clearly demonstrated by the *distinct nature of the diagram* and the provided data. In class B, the score of 80 appears 11 times … whereas, in class A, it has a frequency of only 8. Following the aforementioned steps and procedure, it becomes apparent that class B possesses the largest mode score compared to A, based on the information presented in the *graph*.
R: How do you determine that Company B has the largest salary model among the other companies in question 3?
Hafa: … the determination consists of calculating the frequencies of salary appearances in each company and identifying the value that occurs *most frequently in each case*. By comparing the values and their respective frequencies, it was concluded that Company B has the highest mode salary among the others, specifically 50.
R: What is a mode?
Hafa: A mode refers to the score or value with the highest frequency of occurrence.

Based on the interview, it is apparent that Hafa was able to define mode correctly as "the score or value with the highest frequency of occurrence," yet Hafa used distinct routines to determine the largest mode in items 2 and 3. In item 2, Hafa primarily considered the frequencies of the mode in each class, which was influenced by the shape of the graph, and selected the one with the highest frequency as having the largest mode. However, in item 3, Hafa compared the actual mode values obtained from multiple data groups to identify the class with the largest mode. This signifies that Hafa utilized different procedures when determining the group with the largest mode. The specific routines employed by Hafa are shown in Figure 5.

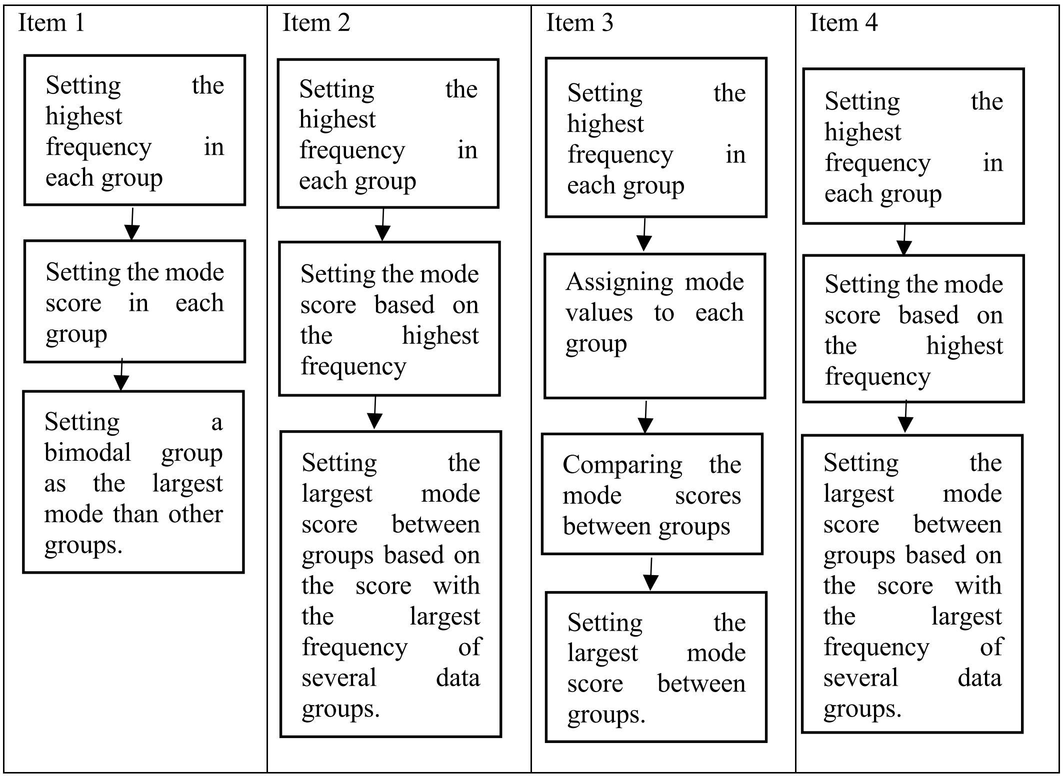


Figure 5. Routines performed by Hafa when engaging in statistical reasoning about the mode

Figure 5 shows the routine used by Hafa in different items. In item 1, the highest frequency within each group was determined, and the mode scores were identified. However, assuming the group had multiple modes, it was selected as having the largest mode. In item 2, Hafa relied solely on the highest frequency among the mode scores of the data groups to determine the largest mode without considering the actual values. In item 3, Hafa compared the mode scores between the data groups to determine the largest mode based on the highest value. In item 4, the group with the highest frequency among the data groups was selected as having the largest mode. These observations suggest that Hafa demonstrated both consistent and inconsistent routines when comparing modes. Table 3 summarizes the directional consistency of her routines across all item pairs, including cases where item 1 involved broader elements than item 2 and item 4. Directional analysis was applied only to Hafa, whose responses varied in the complexity of routine components across specific item pairs. Specifically, item 1 reflected attention to both frequency and the number of modes within a group, whereas items 2 and 4 focused solely on frequency. As a result, consistency was observed from item 1 to items 2 and 4, but not necessarily in the reverse direction.

As shown in Figure 5 and Table 3, Hafa applied a frequency-based procedure when comparing modes in items 1, 2, and 4. Although the approach in item 1 involved recognizing a bimodal pattern, the overall procedure still focused on frequency rather than comparing mode values directly across groups. However, this frequency-based procedure was not appropriate for comparing mode values across multiple groups of data. In contrast, Hafa correctly compared mode values in item 3, but he did not replicate this procedure when answering item 4. This indicates that Hafa applied inconsistent routines when comparing the modes of several data sets. Although one item was answered using an appropriate procedure, she used different and inconsistent procedures for the remaining items. This reflects a lack of procedural consistency in Hafa's statistical reasoning.

Table 3. Routine recapitulation performed by Hafa

Item Number	1	2	3	4
1		consistent*	inconsistent	consistent*
2	inconsistent*		inconsistent	consistent
3	inconsistent	inconsistent		inconsistent
4	inconsistent*	consistent	inconsistent	

Note: Asterisks (*) indicate item pairs where routines in item 1 included multiple focus elements (e.g., frequency and number of modes), whereas items 2 and 4 focused only on frequency. Thus, item 1 → item 2 and item 1 → item 4 are considered consistent, but the reverse directions are not.

4.4. SUMMARY OF FINDINGS ACROSS STUDENTS

To consolidate the analysis of students' routines, correctness, statistical reasoning, routine consistency, and the relationship between their substantiation narratives and endorsed narratives across the four items, Table 4 presents a summary of the findings for all three students. This summary provides a comparative overview of how each student engaged with different data displays, highlighting patterns of consistency, correctness, reasoning, and the extent to which students' justifications aligned with endorsed narratives.

The findings summarized in Table 4 reveal notable variations in the three students' routines, statistical reasoning, routine consistency, substantiation narratives, and their alignment with endorsed narratives across the four items. In particular, in item 2, where the data were presented in dot plots, all students prioritized identifying the highest frequency rather than comparing mode values across groups meaningfully, resulting in incorrect reasoning. Although items 3 and 4 both employed ordered data points, Almi and Hafa exhibited inconsistent routines when comparing modes between similar datasets. These findings suggest that routine consistency does not necessarily ensure accurate statistical reasoning and highlight the occurrence of inconsistent routines even when the data displays remain similar. Additionally, inconsistencies were observed through the substantiation narratives expressed by students when explaining their answers, particularly regarding how they described the data displays and how their routines aligned with the endorsed narratives, revealing a disconnection between the students' routines and the endorsed narratives in several instances.

Table 4. Summary of students' routines and statistical reasoning related to mode across four items

Student	Item Number	Routine Summary	Correct Routine?	Statistical Reasoning	Routine Consistency	Substantiation Narrative and Alignment
Rizi	1	Identify the highest frequency → Determine mode values → Compare mode values → Select the largest mode.	Yes	Focused on comparing mode values across groups in a summary table display.	Inconsistent with Q2; Consistent and Correct (Q3, Q4)	"Class B has the largest *mode value* of 90." (Aligned)
	2	Identify the highest frequency → Select the group with the greatest frequency.	No	Focused on the highest frequency rather than the mode values in a dot plot display.	Inconsistent with Q1, Q3, Q4	"Class B showed the highest mode and represents the *most frequently* occurring data point." (Misaligned)
	3	Identify the highest frequency → Determine mode values → Compare mode values → Select the largest mode.	Yes	Focused on comparing mode values across groups in an ordered data points display.	Inconsistent with Q2; Consistent and Correct (Q1, Q4)	"Company B has a frequency of 4, specifically at a *value of 50.*" (Aligned)
	4	Identify the highest frequency → Determine mode values → Compare mode values → Select the largest mode.	Yes	Focused on comparing mode values across groups in an ordered data points display.	Inconsistent with Q2; Consistent and Correct (Q1, Q3)	"The *mode* in Classes A and B is 60 and *85*; Class B has a mode greater than A." (Aligned)
Almi	1	Identify the highest frequency → Determine mode values → Compare mode values → Select the largest mode.	Yes	Focused on comparing mode values across groups in a summary table display.	Inconsistent with Q2, Q4; Consistent and Correct (Q3)	"Class B has a *mode of 90,* with as many as 10 people." (Aligned)
	2	Identify the highest frequency → Select the group with the greatest frequency.	No	Focused on the highest frequency rather than the mode values in the dot plot display.	Inconsistent with Q1, Q3; Consistent but Incorrect (Q4)	"The mode of Class B is 80, with *a count of 11 individuals*." (Misaligned)
	3	Identify the highest frequency → Determine mode values → Compare mode values → Select the largest mode.	Yes	Focused on comparing mode values across groups in an ordered data points display.	Inconsistent with Q2, Q4; Consistent and Correct (Q1)	"Company B has the *highest mode, a value of 50,* with 4 employees." (Aligned)
	4	Identify the highest frequency → Select the group with the greatest frequency.	No	Focused on the highest frequency rather than mode values in an ordered data points display.	Inconsistent with Q1, Q3; Consistent but Incorrect (Q2)	"Class A has a mode value of 60, with a *frequency of 5* individuals." (Misaligned)

Hafa	1	Identify the highest frequency → Identify mode values → Select the group with multiple mode values.	No	Focused on identifying multiple values with the highest frequency rather than comparing mode values across groups in a summary table display.	Inconsistent with Q3; Consistent but Incorrect (Q2, Q4)	"Class A has the highest frequency, which is 10, derived from *two distinct scores*." (Misaligned)
	2	Identify the highest frequency → Select the group with the greatest frequency.	No	Focused on the highest frequency rather than the mode values in the dot plot display.	Inconsistent with Q1, Q3; Consistent but Incorrect (Q4)	"A score of 80 appeared *11 times in Class B,* whereas only 8 times in Class A." (Misaligned)
	3	Identify the highest frequency → Determine mode values → Compare mode values → Select the largest mode.	Yes	Focused on comparing mode values across groups in an ordered data points display.	Inconsistent with Q1, Q2, Q4	"Company B has a *mode score of 50* with a frequency of 4, which is the highest." (Aligned)
	4	Identify the highest frequency → Select the group with the greatest frequency.	No	Focused on the highest frequency rather than the mode values in an ordered data points display.	Inconsistent with Q1, Q3; Consistent but Incorrect (Q2)	"Class A possesses a larger mode compared to B; classes A and B have mode scores of 60 and 85 with *frequencies of 5* and 3." (Misaligned)

Note: In Table 4, the Routine Summary column represents the original responses provided by the students. The Correct Routine?, Statistical Reasoning, and Routine Consistency columns reflect the researcher's interpretation and evaluation of the students' reasoning processes. The Substantiation Narrative presents the students' original statements, while the Alignment reflects the researcher's evaluation of their correspondence with endorsed narratives

5. DISCUSSION

This study reveals that inconsistencies in students' statistical reasoning about mode might be influenced by two key factors: (a) the way students describe the data display and (b) the disconnection between routine and endorsed narrative. These findings emerged from an investigation into the inconsistency of routines undergraduate students use when performing statistical reasoning about mode across different data displays, including summary tables, dot plots, and ordered data points.

Sfard (2008) introduced commognition as a fusion of communication and cognition aspects, with routines functioning as regulatory mechanisms for word use, visual mediators, and narratives. Rahmatina et al. (2022) extended this framework to statistical discourse, noting that routine is a key component in the commognitive framework that plays an important role in statistical reasoning. In our study, data displayed in summary tables, dot plots, and ordered data points function as visual mediators that might have influenced how students establish and apply routines when performing statistical reasoning about modes. When comparing mode values across multiple data groups, students often selected the group with the highest frequency in the data display as having the largest mode, even when the mode values of different groups were identical. This fixation on visual representations, where students interpreted the highest frequencies as indicating the largest mode, illustrates how students' interactions with visual mediators may have shaped their statistical reasoning. The inconsistent routines identified in our study may be explained through three main themes.

5.1. EVIDENCE OF INCONSISTENT ROUTINES

Our study identified inconsistent routines that undergraduate students used when performing statistical reasoning about mode. Inconsistent routines refer to differences in the procedures students use when comparing modes across multiple data groups in four tasks. The inconsistency in the procedures used to compare modes across different data groups reflected differences in how students reasoned when comparing mode values across groups. Interview findings suggested that students, at times, struggled to compare mode values across groups because they tended to focus on the highest frequency as an indicator of the mode.

Although some students exhibited similar routines in certain tasks, the overall variation in patterns indicates multiple manifestations of inconsistent routines in statistical reasoning. Rizi demonstrated inconsistent routines by correctly comparing mode values in three items (items 1, 3, and 4) but incorrectly focusing on frequencies in one item (item 2). This inconsistency suggests how routines might change even when the core task remains the same. Almi exhibited a different pattern of inconsistency, correctly applying value-based comparisons in two items (1 and 3) while using frequency-based comparisons in two other items (2 and 4). Hafa's case appeared to be the most inconsistent, with correct value-based comparison in only one item (3), while using inappropriate approaches in the other items (1, 2, and 4), including focusing on the number of modes rather than their values. These distinct patterns suggest that inconsistent routines may emerge in structured ways across different problem contexts, revealing challenges in students' statistical reasoning about mode.

These findings, which indicate that students seemingly used inconsistent routines when comparing mode values across tasks, align with Garfield et al.'s (2008) finding that students' statistical reasoning is often inconsistent across different items or topics, depending on the problem context and their experience. In our study, although all tasks involved the same type of statistical problem—comparing mode values across multiple data groups—students still demonstrated inconsistent routines. Our work contributes to the growing recognition that comparing multiple data groups is a fundamental statistical activity (Biehler et al., 2018; Frischemeier, 2019; Shin, 2021) and can serve as a productive tool for motivating learners to engage in statistical reasoning (Ben-Zvi, 2004).

5.2. INFLUENCE OF DATA DISPLAYS ON INCONSISTENT ROUTINES

The type of display used and how the data are represented may determine the trends and predictions that can be made (Jones et al., 2004). This idea builds upon prior research showing that using diagrams is a challenging process requiring substantial experience (van Garderen et al., 2014) and that visual mediators influence an individual's ideas and perceived actions based on the particular choice of visual representation used (Tabach & Nachlieli, 2011). Our findings extend the understanding of how visual representations might shape students' reasoning by demonstrating challenges in describing data, specifically in mode comparison tasks. For instance, when presented with dot plots, all three students incorrectly focused on comparing the frequencies rather than the mode values across data sets. Similarly, in the case of ordered data points, Almi and Hafa correctly compared mode values in one item but reverted to frequency-based comparisons in the other, showing inconsistent use of routines across similar data displays. These specific challenges with visual mediators led students to prioritize the visual aspect of 'tallest shape' or 'highest frequency' over the conceptual understanding of mode as a value. In this regard, it is essential that students distinguish between the concept of a variable and the concept of frequency when comparing modes. This study argues that conceptual understanding of mode serves as a foundation for applying appropriate routines when conducting statistical reasoning in comparing data sets, regardless of the data display format. Garfield (2002) reported that conceptual understanding of key ideas, such as 'center', is a fundamental aspect of statistical reasoning.

Our findings also suggest that undergraduate students' statistical reasoning when describing data displays might exhibit instability when comparing modes across various data displays. The three students in this study employed correct routines in response to one item but failed to do so in their responses to another item with the same type of data display, indicating how data displays might influence the instability of their routines. For instance, Almi and Hafa successfully executed the correct routine in item 3 using ordered data displays but failed to implement the same routine in item 4, which utilized similar data displays. This pattern may suggest that the manner of data presentation influenced the inconsistency in routines adopted by students. Even when confronted with nearly identical data displays, the students applied different routines, indicating that subtle features in data displays may have influenced their routines. This finding aligns with what Lavie and Sfard (2019) described as "the dynamic nature of routines" and challenges traditional perspectives on learning. While Lavie and Sfard (2019) suggested that previous experiences might shape actions in novel situations, our research found that even minor variations in data displays could have led to significant changes in the students' routines, resulting in inconsistent routines despite the statistical context remaining unchanged. Additionally, our study identified cases where students were unable to use their own correct prior experiences when comparing modes across multiple data groups, which may have contributed to the emergence of inconsistent routines.

5.3. DISCONNECT BETWEEN ROUTINE AND ENDORSED NARRATIVE

The present study reveals inconsistencies emerging from the disconnection between routines and endorsed narratives. The definition of mode as 'the value that appears most frequently in a dataset' (Mann, 2013) is accurate for determining the mode within a single dataset. However, when students applied this definition to compare modes across multiple datasets, they demonstrated inconsistent routines, mistakenly focusing on which group had the highest frequency rather than comparing the actual mode values. This disconnect between routines and endorsed narratives suggests that students may verbally understand the concept of mode but fail to apply appropriate routines when the context shifts from single to multiple datasets.

Students developed procedures for determining the largest mode among data groups based solely on the highest frequency, contradicting their own endorsed narrative that mode refers to the score or value with the highest frequency of occurrence. In certain circumstances, students failed to compare the actual values and focused only on comparing frequencies within each group. This disconnection is further evident in the varying substantiation narratives (Mpofu & Pournara, 2018) provided by students when comparing modes across datasets.

The disconnection between routines and endorsed narratives was particularly evident in the case of Rizi, who correctly identified and compared mode values in items 1, 3, and 4, yet failed to apply the same procedure in item 2 despite stating the definition of mode. When interviewed, Rizi could articulate that "a mode is defined as the value that appears most frequently in a given dataset," but still focused on comparing frequencies rather than values when presented with dot plots. This illustrates how the visual representation may have influenced Rizi's routine despite having the correct endorsed narrative.

Kotsopoulos et al. (2009) suggested an important connection between discursive routines and endorsed narratives. Our study revealed that when routines were improperly applied, they could still result in accurate endorsed narratives, creating a disconnect between routines and endorsed narratives, as observed when students compared modes across datasets. Crooks and Alibali (2014) described ‘connection knowledge’ as the understanding of relationships within a domain. In our research, we found a disconnection between students’ routines and their knowledge about the definition of mode when comparing modes across data groups; this was evident in the inconsistency between their procedures and stated definitions. This condition results in inconsistent routines.

These findings have implications for statistical education, particularly when considered alongside Groth and Bergner’s (2006) research which shows that most preservice teachers operate at the unistructural and multistructural levels of thinking. Our findings on inconsistent routines align with Groth and Bergner’s results, as students who focused exclusively on frequency without considering the meaning of mode demonstrated unistructural thinking (focusing on one relevant aspect), while those who recognized both frequency and value but couldn’t consistently integrate them exhibited multistructural thinking (identifying multiple aspects without coherent connections). Because comparing modes across several data groups requires advanced reasoning, routines can serve as valuable tools to help students develop this statistical reasoning, supporting Sfard’s (2020) assertion that routines function as a tool within discourse. Students need consistent practice with various complex problems to develop robust routines for mode comparison tasks.

5.4. IMPLICATIONS AND LIMITATIONS

Statistics teachers and instructors are encouraged to provide teaching materials or exercises related to procedures or steps in comparing the modes of several groups of data, and to provide items with various forms of data display regarding the comparison of the modes of several groups of data. Such efforts might enable students to become accustomed to using the correct and consistent routine in comparing the modes of several groups of data. For stakeholders, students’ skills in using consistent routines correctly when comparing the modes of several data groups can be included in the competencies that students must have in the statistics learning curriculum. This is because one way to evaluate students’ statistical reasoning when comparing modes is by assessing the routines they use for comparison.

This study, however, has several limitations. It was limited to four items, each representing a specific type of quantitative data display (summary tables, graphs, and ordered data points) and involved a relatively small number of undergraduate students (43 in total, with only three interviewed). These factors may restrict the generalizability of the findings. Therefore, further studies are needed to broaden the scope and validate these findings across different contexts.

This research focused on routines in the context of quantitative data and the statistical process of describing data. This study also opens avenues for future research on statistical reasoning at the undergraduate level, particularly in the context of comparing modes in qualitative data (i.e., nominal and ordinal). Future research could expand the number and types of items, involve more participants, and include various forms of data representation to capture a broader range of reasoning patterns. Further investigation could examine how students apply routines in the processes of organizing, representing, analyzing, and interpreting categorical data, and whether these routines reflect conceptual understanding or procedural tendencies. Additional research could also explore students’ routines at the secondary school level to determine whether consistent routines reflect deeper conceptual understanding or procedural habit when engaging in statistical reasoning.

6. CONCLUSION

Statistical reasoning is crucial in comparing modes across multiple datasets. Fostering this skill may be supported by exposing students to various data displays, thus allowing them to develop experiences in statistical reasoning concerning mode. Different contexts and data displays might influence the routines adopted by students when comparing modes across multiple datasets. Furthermore, comparing the modes of several data groups from various data displays may involve complex statistical reasoning. Therefore, routines are essential in supporting statistical reasoning.

ACKNOWLEDGEMENTS

The authors would like to extend their deepest appreciation to Maxine Pfannkuch, whose scholarly insights and constructive feedback were instrumental in the reorganization and refinement of this manuscript. Her guidance significantly contributed to the clarity and coherence of the manuscript. We also wish to thank the reviewers for valuable comments and suggestions, which helped improve the quality of this manuscript, as well as the students who participated in this study.

REFERENCES

Bakker, A. (2003). The early history of average values and implications for education. *Journal of Statistics Education*, *11*(1). https://doi.org/10.1080/10691898.2003.11910694

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D., A. (2020). *Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II): A framework for statistics and data science education*. American Statistical Association and National Council of Teachers of Mathematics.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, *3*(2), 42–63. https://doi.org/10.52041/serj.v3i2.547

Bennett, J., Briggs, W. L., & Triola, M. F. (2017). *Statistical reasoning for everyday life* (5th ed.). Pearson Education Limited.

Biehler, R., Frischemeier, D., Reading, C., & Shaughnessy, J. M. (2018). Reasoning about data. In D. Ben-Zvi, K. Makar, & J. Garfield, (Eds.), *International handbook of research in statistics education* (pp. 139–192). Springer. https://doi.org/10.1007/978-3-319-66195-7_5

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, *34*(4), 344–377. https://doi.org/10.1016/j.dr.2014.10.001

Fernández-León, A., Gavilán-Izquierdo, J. M., González-Regaña, A. J., Martín-Molina, V., & Toscano, R. (2021). Identifying routines in the discourse of undergraduate students when defining. *Mathematics Education Research Journal*, *33*(2), 301–319. https://doi.org/10.1007/s13394-019-00301-1

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158. https://doi.org/10.2307/749671

Frischemeier, D. (2019). Statistical reasoning when comparing groups with software—Frameworks and their application to qualitative video data. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research: International perspectives* (pp. 285–305). Springer. https://doi.org/10.1007/978-3-030-03472-6_13

Garfield, J., & Chance, B. (2000). Assessment in Statistics Education: Issues and Challenges. *Mathematical Thinking and Learning*, *2*(1–2), 99–125. http://dx.doi.org/10.1207/S15327833MTL0202_5

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910676

Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). *Developing students' statistical reasoning*. Springer. https://doi.org/10.1007/978-1-4020-8383-9

Groth, R. E., & Bergner, & J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, *8*(1), 37–63. https://doi.org/10.1207/s15327833mtl0801_3

Güçler, B. (2013). Examining the discourse on the limit concept in a beginning-level calculus classroom. *Educational Studies in Mathematics*, *82*(3), 439–453. https://doi.org/10.1007/s10649-012-9438-2

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Erlbaum Associates.

Ioannou, M. (2018). Commognitive analysis of undergraduate mathematics students' first encounter

with the subgroup test. *Mathematics Education Research Journal*, *30*(2), 117–142. https://doi.org/10.1007/s13394-017-0222-6
Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics*, *96*(1), 1–16. https://doi.org/10.1007/s10649-017-9761-8
Jones, G. A., Langrall, C. W., Mooney, E. S., & Thornton, C. A. (2004). Models of development in statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 97–117). Springer. https://doi.org/10.1007/1-4020-2278-6_5
Karatoprak, R., Karagöz Akar, G., & Börkan, B. (2015). Prospective elementary and secondary school mathematics teachers' statistical reasoning. *International Electronic Journal of Elementary Education*, *7*(2), 107–124. https://iejee.com/index.php/IEJEE/article/view/69/67
Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, *24*(5), 392–414. https://doi.org/10.5951/jresematheduc.24.5.0392
Kotsopoulos, D., Lee, J., Heide, D., & Schell, A. (2009). Discursive routines and endorsed narratives as instances of mathematical cognition. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the 31st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 42–49). Georgia State University. https://pmena.org/pmenaproceedings/PMENA%2031%202009%20Proceedings.pdf
Landtblom, K., & Sumpter, L. (2019). Teachers and prospective teachers' conceptions about averages. *Journal of Adult Learning, Knowledge and Innovation*, *4*(1), 1–8. https://doi.org/10.1556/2059.03.2019.02
Landtblom, K. (2023). Opportunities to learn mean, median, and mode afforded by textbook tasks. *Statistics Education Research Journal*, *22*(3), Article 6. https://doi.org/10.52041/serj.v22i3.655
Lavie, I., & Sfard, A. (2019). How children individualize numerical routines: Elements of a discursive theory in making. *Journal of the Learning Sciences*, *28*(4–5), 419–461. https://doi.org/10.1080/10508406.2019.1646650
Leavy, A., & O'Loughlin, N. (2006). Preservice teachers' understanding of the mean: Moving beyond the arithmetic average. *Journal of Mathematics Teacher Education*, *9*(1), 53–90. https://doi.org/10.1007/s10857-006-9003-y
Magaldi, D., & Berler, M. (2020). Semi-structured interviews. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 4825–4830). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-319-24612-3
Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 353–373). Springer. https://doi.org/10.1007/1-4020-2278-6_15
Mann, P. S. (2013). *Introductory statistics* (8th ed.). John Wiley & Sons.
Martin, W. G., Carter, J., Forster, S., Howe, R., Kader, G., Kepner, H., Quander, J. R., McCallum, W., Robinson, E., Sniper, V., & Valdez, P. (2009). *Focus in high school mathematics: Reasoning and sense making*. National Council of Teachers of Mathematics.
Mpofu, S., & Pournara, C. (2018). Learner participation in the functions discourse: A focus on asymptotes of the hyperbola. *African Journal of Research in Mathematics, Science and Technology Education*, *22*(1), 2–13. https://doi.org/10.1080/18117295.2017.1409170
Nardi, E., Ryve, A., Stadler, E., & Viirman, O. (2014). Commognitive analyses of the learning and teaching of mathematics at university level: The case of discursive shifts in the study of calculus. *Research in Mathematics Education*, *16*(2), 182–198. https://doi.org/10.1080/14794802.2014.918338
Ng, O. L. (2016). The interplay between language, gestures, dragging and diagrams in bilingual learners' mathematical communications. *Educational Studies in Mathematics*, *91*(3), 307–326. https://doi.org/10.1007/s10649-015-9652-9
Park, M.-S., & Lee, K.-H. (2014). The impact of a teacher's attention deriving on students' statistical discourse. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014). Flagstaff, Arizona, USA.* International Statistical Institute. https://icots.info/9/proceedings/pdfs/ICOTS9_3B3_PARK.pdf

Pratiwi, E., Nusantara, T., Susiswo, S., & Muksar, M. (2022). Routines' errors when solving mathematics problems cause cognitive conflict. *International Journal of Evaluation and Research in Education*, *11*(2), 773–779. https://doi.org/10.11591/ijere.v11i2.21911

Rahmatina, D., Nusantara, T., Prata, I. N., & Susanto, H. (2022). Statistical reasoning process of students in decision making using commognitive framework. *Acta Scientiae*, *24*(3), 63–88. https://doi.org/10.17648/acta.scientiae.6603

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. https://doi.org/10.1037/a0014224

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge University Press.

Sfard, A. (2018). Commognition. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 1–7). Springer. https://doi.org/10.1007/978-3-319-77487-9_100031-1

Sfard A. (2020). Commognition. In S. Lerman (Ed.). *Encyclopedia of mathematics education* (pp. 95–101). Springer. https://doi.org/10.1007/978-3-030-15789-0_100031

Shin, D. (2021). A framework for understanding how preservice teachers notice students' statistical reasoning about comparing groups. *International Journal of Mathematical Education in Science and Technology*, *52*(5), 699–720. https://doi.org/10.1080/0020739X.2019.1699968

Tabach, M., & Nachlieli, T. (2011). Combining theories to analyse classroom discourse: A method to study learning process. In M. Pytlak, R. Rowland, & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education. CERME 7* (pp. 2524–2532). University of Rzeszów, Poland. https://wwwold.mathematik.tu-dortmund.de/~prediger/ERME/CERME7-Proceedings-2011.pdf

Tabach, M., & Nachlieli, T. (2015). Classroom engagement towards using definitions for developing mathematical objects: The case of function. *Educational Studies in Mathematics*, *90*(2), 163–187. https://doi.org/10.1007/s10649-015-9624-0

Tabach, M., & Nachlieli, T. (2016). Communicational perspectives on learning and teaching mathematics: prologue. *Educational Studies in Mathematics*, *91*(3), 299–306. https://doi.org/10.1007/s10649-015-9638-7

Toscano, R., Gavilán-Izquierdo, J. M., & Sánchez, V. (2019). A study of pre-service primary teachers' discourse when solving didactic-mathematical tasks. *Eurasia Journal of Mathematics, Science and Technology Education*, *15*(11), 1–16. https://doi.org/10.29333/ejmste/108631

Turner, P., & Turner, S. (2009). Triangulation in practice. *Virtual Reality*, *13*(3), 171–181. https://doi.org/10.1007/s10055-009-0117-2

van Garderen, D., Scheuermann, A., & Poch, A. (2014). Challenges students identified with a learning disability and as high-achieving experience when using diagrams as a visualization tool to solve mathematics word problems. *ZDM Mathematics Education*, *46*(1), 135–149. https://doi.org/10.1007/s11858-013-0519-1

Viirman, O. (2015). Explanation, motivation and question posing routines in university mathematics teachers' pedagogical discourse: A commognitive analysis. *International Journal of Mathematical Education in Science and Technology*, *46*(8), 1165–1181. https://doi.org/10.1080/0020739X.2015.1034206

Wang, S., & Kinzel, M. (2014). How do they know it is a parallelogram? Analysing geometric discourse at van Hiele level 3. *Research in Mathematics Education*, *16*(3), 288–305. https://doi.org/10.1080/14794802.2014.933711

Zayyadi, M., Nusantara, T., Subanji, Hidayanto, E., & Sulandra, I. M. (2019). A commognitive framework: The process of solving mathematical problems of middle school students. *International Journal of Learning, Teaching and Educational Research*, *18*(2), 89–102. https://doi.org/10.26803/ijlter.18.2.7

DESI RAHMATINA
Faculty of Teacher Training and Education
Universitas Maritim Raja Ali Haji
Indonesia