Statistical reasoning levels of high school students in solving statistics related problems

R Setianingsih, DA Rahmah

Department of Mathematics, Universitas Negeri Surabaya, Indonesia. Gedung C8, Kampus Unesa Ketintang, Surabaya, Indonesia 60231

Corresponding author: rinisetianingsih@unesa.ac.id

Abstract. Statistical reasoning is one of the focus of statistical learning at this time. Statistical reasoning is needed by students to understand the concept of statistics, to process information, and to draw conclusions both in statistical learning and in daily life. This research is a descriptive study with a qualitative approach that aims to describe the level of statistical reasoning of high school students who have low mathematical abilities in solving statistics related problems in each statistical process, namely the process of describing data, the process of organizing and reducing data, the process of representing data, as well as the process of analyzing and interpreting data. The subjects in this study were two high school students of grade XII who had low mathematical abilities based on the math skills tests that had been given. The results showed that both subjects in the process of describing data successively reached level 2 (Verbal Reasoning) and level 4 (Procedural Reasoning). Meanwhile, in the process of organizing and reducing data and the process of representing data, the first and second subjects both reached level 1 (Idiosyncratic Reasoning), while in the process of analyzing and interpreting data, the first and second subjects both reached level 1 (Idiosyncratic Reasoning).

Keywords: Statistics, statistical reasoning, level of statistical reasoning.

1. Introduction
Statistics is a scientific discipline dealing with data collection, data processing, data presentation, and decision making. Statistics is knowledge about ways of collecting data, processing or analyzing data, and drawing conclusions based on data analysis [1]. In addition, statistics is a tool that can be used to solve problems that occur in daily life, both at work, and in science [2]. There are four statistical processes that a person can go through in understanding statistical information and concepts, namely the process of describing data, organizing and reducing data, representing data, as well as analyzing and interpreting data [3].

Taking into account the important role of statistics in daily life, the development of statistical learning is always ongoing. In the last decades, the focus of statistical learning has shifted from only carrying out calculations with the right formulas to understanding statistical concepts and developing statistical reasoning [4]. These changes occurred, because the assessment of student learning outcomes through daily tests that previously emphasized the calculation results were felt to be unable to provide a clear picture of student understanding related to statistical concepts.
In addition, to develop statistical reasoning, it is important for teachers to pay attention to the factors underlying achievement of statistical learning [5]. It is said that one of the factors that underlies the success of students in statistical learning is their mathematical abilities, namely basic computing skills and algebra [6]. This opinion is supported by the results of study which revealed that students who scored 7 or less on the basic mathematics quiz had significantly got statistical scores less than two than students who scored 10 or more [7]. This shows that students with different levels of mathematical ability will have different statistical abilities, which can further influence the reasoning and statistical thinking that students can do.

Statistical reasoning can be defined as the ability to explain why and how an answer or result is obtained and how to draw a conclusion [8]. Then it also stated that statistical reasoning is a way of thinking using statistical information and concepts [9]. This opinion is supported by the views that statistical reasoning is a person's ability to understand and relate statistical concepts to interpret data and draw conclusions based on a given context [10]. Meanwhile, what is meant by statistical reasoning in this research is a logical thinking process in order to understand data and statistical information through the process of describing data, organizing and reducing data, representing data, as well as analyzing and interpreting data.

In this study, the researchers used a statistical reasoning model of Garfield (2002) to describe the statistical reasoning profile of high school students in solving statistical problems, which consists of five levels of reasoning, namely, Idiosyncratic Reasoning, Verbal Reasoning, Transitional Reasoning, Procedural Reasoning, and Integrated Reasoning. This is due to the reason that the statistical reasoning model introduced by Garfield refers to statistical material that students can understand [11]. Therefore, by using this model, the researchers can place students at a level of statistical reasoning in accordance with students' understanding of statistical material.

2. Method
This research is a descriptive study with a qualitative approach that aims to describe the level of statistical reasoning of high school students who have low mathematical abilities in solving statistical problems in each statistical process. The steps to obtain data in this study are as follows: First, to determine the subject of research, the researchers conducted mathematics ability tests to students of grade XII MIA programs who have learned statistical material in the previous lesson.

Based on the math ability test scores, two students were selected who had the lowest math ability test scores. Second, to obtain the data of students' level of statistical reasoning in solving statistical problems, the researchers conducted statistical reasoning tests related to the process of describing data, the process of organizing and reducing data, the process of representing data, as well as the process of analyzing and interpreting data.

Finally, the researchers conducted an interview with the two research subjects after the statistical reasoning test was completed. The interview aims to reveal more deeply about the students' reasoning that is not visible from the answers to the statistical reasoning tests. After the data from the statistical reasoning test and interview process were obtained, the researchers then analyzed the data in order to describe the level of statistical reasoning of high school students in solving statistical problems in each statistical process.

3. Results and Discussion
Based on the results of the statistical reasoning tests and the interview with the research subjects, data were obtained regarding the level of statistical reasoning of students who have low mathematical abilities in solving statistical problems in each statistical process as follows:

a. Level of Student Statistical Reasoning in the Process of Describing Data
The first subject is able to write three different descriptions, but of the three descriptions written by the first subject, only one description that fits the information table which was given. In addition, the description given by the first subject during the interview process was only partly correct. This is the
reason used by the researchers to put the first subject in the process of describing data into the second level of statistical reasoning (Verbal Reasoning).

As with the first subject, the second subject was able to write three different descriptions, but the three descriptions written by the second subject did not match the information table provided. During the interview process, the second subject was able to provide three descriptions and only one description had an incomplete explanation. Therefore, the researchers placed the second subject's level of statistical reasoning during the process of describing data at level 4 (Procedural Reasoning).

b. **Level of Statistical Reasoning in the Process of Organizing and Reducing Data**

In the process of organizing and reducing data, the first and second subjects were asked to determine the age of the driver involved in most motorcycle accidents (driver age mode) in Britain, in 2016. However, before determining the age mode of the driver, the first subject and second subject must first calculate the number of drivers involved in motorcycle accidents in Britain in 2016, for each age interval by using data on the number of drivers involved in motorcycle accidents in Britain, in 2017, and data on the percentage change in the number of drivers involved in accidents since 2016.

The results of the statistical reasoning test and the interview process show that the two research subjects could not count the number of drivers involved in motorcycle accidents in Britain in 2016 for each age interval. Both research subjects determine the age mode of drivers involved in accidents in Britain in 2016 based solely on data on the number of drivers involved in motorcycle accidents in Britain in 2017 (without calculating the data in 2016), so that the answers of the two research subjects were incorrect. Therefore, researchers put the level of statistical reasoning on the first subject and second subject during the process of organizing and reducing data at level 1 (idiosyncratic reasoning). The following are the written answers as well as the results of researchers' interviews with the first subject and the second subject in the process of organizing and reducing data.

Table 1: Answers to Statistical Reasoning Tests and Interview Transcript in the Process of Organizing and Reducing Data

| Subject 1 | Subject 2 |
|-----------|-----------|
| 27. Pada usia 20-29, yaitu berjumlah 5.986 orang, percentage -10.5% | 27. Pada usia 30-39 |
| R: Determine the age of the drivers involved in most accidents in Britain in 2016. Well, what was the purpose of that question you were asked? | R: Determine the age of drivers involved in most accidents in Britain in 2016. If the problem is like that, then which data do you need? |
| S: Determine the age of the motorcycle driver involved in the accident since 2016. | S: (While pointing to the table on the question sheet), in motorcycle column, the number of drivers involved in the accident and the percentage change in the number of drivers involved in the accident since 2016. |
| R: Then, the data from the table that you need means which column data? | R: Then, how about the first step you take to answer question number 2? |
| S: (not immediately answered) the number of drivers in 2017 with the number of drivers injured. | S: I did not write the way, Sis. |
| R: Then, how was the first step you took to be able to determine the age involved in most accidents in 2016? | R: Then how do you determine it? |
| S: (Be quiet, don't answer right away) directly, Sis. |
| R: What do you think? | S: Because, the decrease percentage is lower at the age of 30-39. |
| S: At the age of 20 to 29 that is 5986 with a percentage of -10.5%. |
c. Level of Statistical Reasoning in the Process of Representing Data

In the process of representing data, the first subject and the second subject were asked to make a diagram or graph that can make it easier for readers to compare two groups of data, each group consisting of data intervals of age and frequency. The two research subjects each draw two line-diagrams on the answer sheet of the statistical reasoning test, and explained how they constructed them in the interview process. Based on these answers, it appears that the answers of the two research subjects do not match the imperative sentence on the test of statistical reasoning. In addition, the line diagrams made by the two subjects are not appropriate to illustrate data in the form of class intervals of the driver's age and the number of drivers involved in the accident for each age class interval.

It is important note that the graph used to display data in the form of certain classes along with the frequency of each class is a histogram. Meanwhile, line charts are usually used to present continuous data [12]. Thus, the line diagrams made by the two research subjects are not suitable to represent the data provided, so the researchers place the level of statistical reasoning of the first subject and the second subject during the process of representing data at level 1 (Idiosyncratic Reasoning).

d. Level of Statistical Reasoning in the Process of Analyzing and Interpreting Data

In the process of analyzing and interpreting data, the first and second subjects were asked to give a "agree" or "disagree" response to the statement that a car accident is more at risk of making the driver hurt than a motorcycle accident. Both research subjects both responded "agreed" to the statement on the grounds that based on data in the table, the number of drivers injured in a car accident was more than in a motorcycle accident. The way to determine the risk of injury from an accident should be to calculate the average number of drivers injured by the number of drivers involved in the accident. Meanwhile, the answers to the statistical reasoning test and the interview process showed that the two research subjects did not pay attention to data on the number of drivers involved in car and motorcycle accidents. Thus, the "agreed" responses given by the two research subjects are not appropriate, so the researcher places the level of statistical reasoning on the first subject and the second subject during the process of analyzing and interpreting data at level 1 (Idiosyncratic Reasoning).

4. Conclusion

The results showed that students who have low mathematical abilities are at level 2 (Verbal Reasoning) and level 4 (Procedural Reasoning) in the process of describing data and are at level 1 (Idiosyncratic Reasoning) in the process of organizing and reducing data, in the process of representing data, as well as in the process of analyzing and interpreting data. In addition, in the process of organizing and reducing data, from the beginning students cannot use the percentage rules to obtain new data, so that the students cannot reduce the data using a measure of central tendency (mode) for the new data. In the process of analyzing and interpreting data, students with low mathematical ability cannot draw conclusions correctly according to the context of the data provided. That is because students do not really understand the concept of comparison.

Based on these results it can be concluded that, the cause of the low level of statistical reasoning of students in conducting statistical processes is the low understanding of the mathematical concepts that underlie these statistical concepts. This is also in line with the results of study which revealed that
students who scored 7 or less on the basic mathematics quiz had significantly got statistical scores less than two than students who scored 10 or more [7]. This shows that students with different levels of mathematical ability will have different statistical abilities, which can further influence the reasoning and statistical thinking that students can do. Therefore, it is important for teachers at the beginning of learning statistics to ensure that students have sufficient mathematical knowledge as a provision to learn statistical concepts.

References
[1] Sudjana. 2009. Metode Statistika. Bandung: Tarsito.
[2] Moore, D.S. 1997. “New Pedagogy and New Content: The Case of Statistics”. International Statistics Review. Vol. 65(2): pp. 123–165.
[3] Jones, A.G., Thornton, C.A, Langrall, C.W., Mooney, E.S., Perry, B., and Putt, I.J.A. 2000. “Framework for characterizing children’s statistical thinking”. Math Think Learn. Vol. 2: pp. 269–307. DOI: 10.1207/S15327833MTL0204_3.
[4] Mooney, E.S. 2002. “A framework for characterizing middle school students’ statistical thinking”. Math Think Learn. Vol. 4: pp. 23–63. DOI: 10.1207/S1 532 7833MTL0401_2.
[5] Qian, G. 2011. “Teaching, Learning and Retention of Statistical Ideas in Introductory Statistics Education”. European Journal of Pure and Applied Mathematics. Vol. 4 (2): pp. 103–116.
[6] Galli, S., Chiesi, F., and Primi, C. 2010. “Assessing Mathematics Competence in Introductory Statistics Course: An Application of the Item Response Theory”. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia. Voorburg, the Netherlands: International Statistical Institute, (Online), (www.stat.auckland.ac.nz/~iase/publications.php, downloaded on 4 February 2019).
[7] Johnson, M. and Kuennen, E. 2006. “Basic Math Skill and Performance in an Introductory Statistics Course”. Journal of Statistics Education. Vol. 14: pp. 2. DOI: 10.1080/10691898.2006.11910581.
[8] DelMas, R. 2002. “Statistical Literacy, Reasoning, and Learning: A Commentary”. Journal of Statistics Education. Vol. 10(3). DOI: 10.1080/10691898.2002.11910679.
[9] Ben-Zvi, D. and Gafield, J. 2004. The Challenge of Developing Statistical Literacy, Reasoning, and Thinking. Boston MA: Kluwer Academic Publisher.
[10] Olani, A., Hoekstra, R., Harskamp, E., and van der Welf, G. 2011. “Statistical Reasoning Ability, Self-Efficacy, and Value Beliefs in a Reform Based University Statistics Course”. Electronic Journal of Research in Educational Psychology, (Online), Vol. 9(1): pp. 49–72, (https://www.researchgate.net/publication/287486979_Statistical_Reasoning_Ability_SelfEfficacy_andValue_Beliefs_in_a_Reform_Based_University_Statistics_Course, downloaded on 8 October, 2018).
[11] Yusuf, Y. 2017. “Konstruksi Penalaran Statistis pada Statistika Penelitian”. Scholaria. Vol. 7(1): pp. 60–69.
[12] As’ari, A.R., Chandra, T.D., Yuwono, I., Anwar, L., Nasution, S.H., Hasanah, D., Muksar, M., Sari, V.K., dan Atikah, N. 2018. Matematika. Jakarta: Kemendikbud.